ISSN 2328-2231 (Print)

From Knowledge to Wisdom

Journal of Control Science and Engineering

Volume1, Number1, December 2013

DAVID PUBLISHING
David Publishing Company
www.davidpublishing.com

Journal of Control Science and Engineering
Volume 1, Number 1, December 2013 (Serial Number 1)

Journal of Control Science and Engineering

Volume 1, Number 1, December 2013 (Serial Number 1)

DAVID PUBLISHING

David Publishing Company
www.davidpublishing.com

Publication Information:
Journal of Control Science and Engineering is published monthly in hard copy (ISSN 2328-2231) by David Publishing Company located at 240 Nagle Avenue #15C, New York, NY 10034, USA.

Aims and Scope:
Journal of Control Science and Engineering, a monthly professional academic journal, covers all sorts of papers which illustrate the application of control theory and control techniques, and its supporting tools in all possible areas of automation.

Manuscripts and correspondence are invited for publication. You can submit your papers via Web Submission, or E-mail to control@davidpublishing.org or control@davidpublishing.com. Submission guidelines and Web Submission system are available at http://www.davidpublishing.com.

Editorial Office:
240 Nagle Avenue #15C, New York, NY 10034, USA
Tel: 1-323-984-7526, 323-410-1082; Fax: 1-323-984-7374, 323-908-0457
E-mail: control@davidpublishing.org; control@davidpublishing.com.

Abstracted / Indexed in:
ProQuest
Google Scholar
Ulrich's Periodicals Directory
Summon Serials Solutions
CNKI
Chinese Scientific Journals Database, VIP Corporation, Chongqing, P.R.C.

Subscription Information:
Price (per year):
Print $520/Online $360
Print and Online $680

David Publishing Company
240 Nagle Avenue #15C, New York, NY 10034, USA
Tel: 1-323-984-7526, 323-410-1082; Fax: 1-323-984-7374, 323-908-0457
E-mail: order@davidpublishing.com

DAVID PUBLISHING

David Publishing Company
www.davidpublishing.com

Journal of Control Science and Engineering

Volume 1, Number 1, December 2013 (Serial Number 1)

Contents

Journal of Control Science and Engineering 1 (2013) 1-12

Hardware Type 2 Fuzzy Logic Position Controller Based on Karnik-Mendel Algorithms

Pedro Ponce-Cruz, Arturo Molina and Arturo Tellez-Velazquez
Tecnologico de Monterrey, Distrito Federal 07240, Mexico

Abstract: This paper presents an analysis of the KM (Karnik-Mendel) algorithms performance under real time implementation using 3 types: the non-iterative, the iterative and the enhanced, and their feasibility for real-time interval type 2 fuzzy logic control system applications. The results are also compared against NT (Nie-Tan) method that is one of the fastest and simplest defuzzification methods. Because the DC (direct current) servo-motor is one of the most used motors in different industrial applications and the model of the motor is nonlinear, this motor was selected for validating the implementation in real time hardware. This DC motor is a perfect option for studying the real time performance of KM algorithms in order to show up its limits and possibilities for real-time control system applications. These methodologies are implemented in National Instruments LabVIEW FPGA (field programmable gate array) module hardware which is one of the most used platforms in the industry. The results show that the E-KM (enhanced KM) algorithm and the NT method present good results for implementing real-time control applications in real time hardware. Although fuzzy logic type 2 is a good option for working with nonlinear and noise from the sensors, the defuzzification method has to react in a short period of time in order to allow good control response. Hence, a complete study of defuzzification is needed for improving the real time implementations of fuzzy type 2.

Key words: Fuzzy logic type 2, KM algorithms, NT method, defuzzification, type-reduction, DC servo-motor control.

1 .Introduction

Different research works present three KM (Karnik-Mendel) algorithms for the IT2FLS (interval type 2 fuzzy logic systems): the non-iterative [1] and the iterative [2] types that provide the same numerical results; the main difference is the improved total iteration count of the iterative type compared with the non-iterative one. Actually, both types are outperformed by the enhanced KM algorithm [3] which needs several initial conditions and lets the system converge faster than the other ones. These defuzzification methods get the generalized centroid of an IT2FLS.

Different applications require opportune decisions; those decisions which require the noise immunity that the IT2FLS can provide [4]. The KM algorithms are very intensive and sometimes they are not appropriated for real-time applications. When they are dealing with a big amount of data, a late response appears generating an incorrect decision which cannot be acceptable in real-time hardware systems.

The hardware implementation of IT2FLS offers good results but the software implementation is not good enough as it was shown in Ref. [5]. A hardware implementation is considered in this work for proving the KM algorithms in real-time applications.

For this purpose, a DC (direct current) servo-motor is analyzed. The IT2FLS is implemented in hardware by FPGA (field programmable gate array) based on LabVIEW and each method was implemented and tested independently. Two hardware considerations were taken: the first one is the swiftness and the second one is the final resource utilization. Every algorithm was compared using these two hardware conditions; also this paper shows a complete chart that

Corresponding author: Pedro Ponce-Cruz, Dr., research fields: robotics, control systems, instrumentation and artificial intelligence. E-mail: pedro.ponce@itesm.mx.

presents the complexity of every algorithm according to the discrete discourse universe points.

This paper defines a new procedure of establishing the hardware implementation based on LabVIEW FPGA for KM algorithms. According to the results presented, various KM algorithms are not appropriated for real-time control applications. Finally, expert must take into account the ratio between speed and area.

2. Interval Type 2 Fuzzy Logic Systems

The IT2FLS topology is the same as the T1FLS (type 1 fuzzy logic systems). It provides the fuzzification, the inference and the defuzzification stages.

The IT2FLS fuzzification maps the crisp values x_j into several membership values according to its membership degree. This means that a crisp value could belong to more than one IT2FS (interval type 2 fuzzy sets); those two membership degrees form a FOU (footprint of uncertainty). Each membership obtained after defuzzification process is related using fuzzy logic operations as conjunction (AND) and disjunction (OR). The fuzzy conjunction allows expert relating all the implied premises (input sets) and the fuzzy disjunction allows expert aggregating all these implied values in order to obtain a specific consequent (output set). With all these relations, a rule set is built. The rule set represents the IT2FLS conventional configuration.

Fig. 1 shows a complete picture of the IT2FLS which can be used in hardware implementations. Each block can be an independent hardware entity. The inferred set is calculated and stored in a memory location.

3. The KM Algorithms

In previous works [1-3, 6], the KM algorithms includes several modifications in order to decrement the impact of the large number of iterations required to generate a single centroid. This method is computationally exhaustive, different research works proposed some initial conditions and modifications based on the seminal algorithm. These modifications have the purpose of decreasing the search space for both the left and the right centroids.

Assume that the output discourse universe Y is a set of all the possible crisp outputs that can be obtained from a defuzzification method, i.e. $y \in Y$. Now, let B be a consequent set defined along Y. The key of finding each approximated centroid is to find a switch point where the ES (embedded set) will change from one outer membership function to another; specifically, the LMF (lower membership function) or $\underline{\mu}_{\tilde{B}}(y)$ and

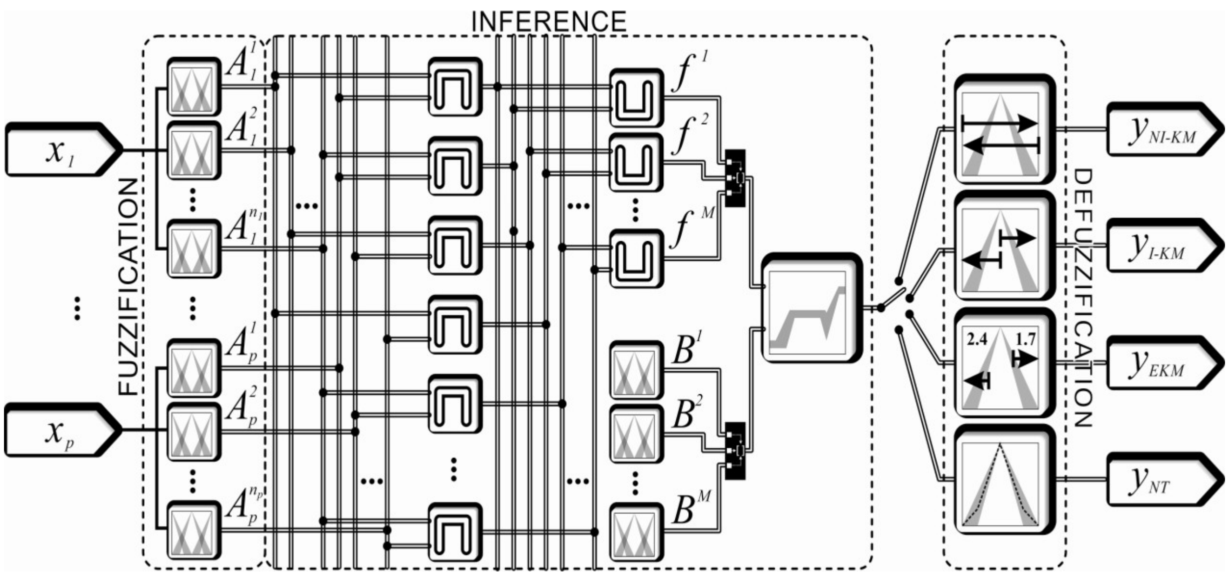


Fig. 1 IT2FLS with four defuzzification methods: the non-iterative, iterative and enhanced KM algorithms, and the NT (Nie-Tan) method.

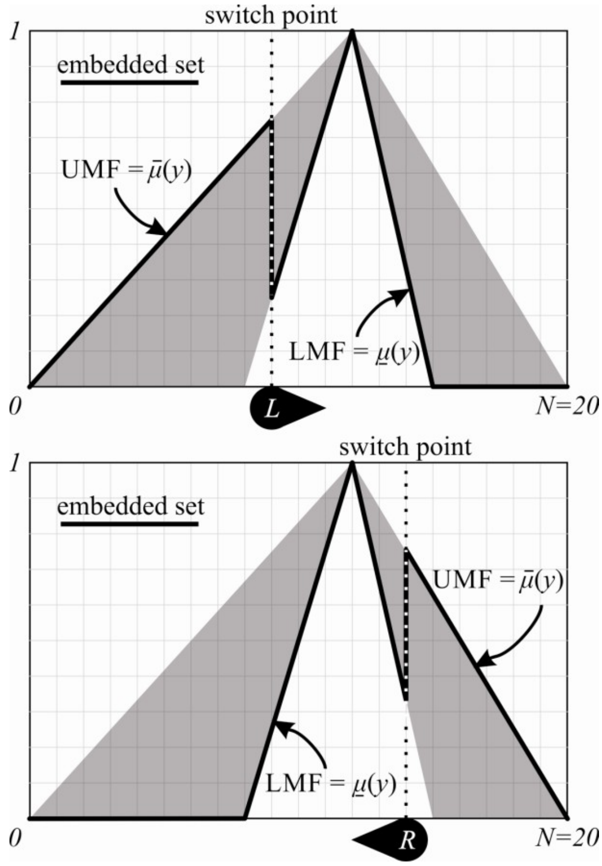


Fig. 2 The general centroid search in the KM algorithms. There is not centroid search for this algorithm; all the centroids are computed. This method is known as the NI-KM (non-iterative Karnik-Mendel) algorithm.

the UMF (upper membership function) or $\bar{\mu}_{\tilde{B}}(y)$, which comprises the FOU. Fig. 2 provides a graphical description of the KM algorithm, in general.

3.1 Non-Iterative Version

The NI-KM algorithm [1] creates all the embedded sets θ_L and θ_R as possible from 1 to N. For finding the left and right centroids, the expert must calculate the following formulas:

$$\theta_L = \begin{cases} \bar{\mu}_{\tilde{B}}(y_i) & y_i \le L \\ \underline{\mu}_{\tilde{B}}(y_i) & \text{otherwise} \end{cases}$$

$$c_{\theta L} = \frac{\sum_{i=1}^{N} y_i \cdot \mu_{\theta L}(y_i)}{\sum_{i=1}^{N} \mu_{\theta L}(y_i)} \tag{1}$$

where, $c_{\theta L} \in \Theta_l$

$c_l \in \min(\Theta_l)$

$$\theta_R = \begin{cases} \underline{\mu}_{\tilde{B}}(y_i) & y_i \le R \\ \bar{\mu}_{\tilde{B}}(y_i) & \text{otherwise} \end{cases}$$

$$c_{\theta R} = \frac{\sum_{i=1}^{N} y_i \cdot \mu_{\theta R}(y_i)}{\sum_{i=1}^{N} \mu_{\theta R}(y_i)} \tag{2}$$

where, $c_{\theta R} \in \Theta_r$

$c_r \in \max(\Theta_r)$.

This non-iterative version creates a vector of left centroids Θ_l and a vector of right centroids Θ_r; from the left centroid vector, the far left centroid (the minimum) is selected while from the right centroid (maximum) vector, the far right centroid is selected. Fig. 3 shows the non-iterative version.

Eqs. (1) and (2) can be rewritten as two summations in the numerator and the denominator as follow:

$$c_{\theta L} = \frac{\sum_{i=1}^{L} y_i \cdot \bar{\mu}_{\tilde{B}}(y_i) + \sum_{i=L+1}^{N} y_i \cdot \underline{\mu}_{\tilde{B}}(y_i)}{\sum_{i=1}^{L} \bar{\mu}_{\tilde{B}}(y_i) + \sum_{i=L+1}^{N} \underline{\mu}_{\tilde{B}}(y_i)} \tag{3}$$

where, $c_{\theta L} \in \Theta_l$

$c_l \in \min(\Theta_l)$

$$c_{\theta R} = \frac{\sum_{i=1}^{R} y_i \cdot \underline{\mu}_{\tilde{B}}(y_i) + \sum_{i=R+1}^{N} y_i \cdot \bar{\mu}_{\tilde{B}}(y_i)}{\sum_{i=1}^{R} \underline{\mu}_{\tilde{B}}(y_i) + \sum_{i=R+1}^{N} \bar{\mu}_{\tilde{B}}(y_i)} \tag{4}$$

where, $c_{\theta R} \in \Theta_r$

$c_r \in \max(\Theta_r)$.

3.2 Iterative Version

This iterative version [2, 7, 8] searches for the left and right centroids starting from a convenient initial embedded set θ_i. After the initial centroid, the following centroid search of θ_i for the left and right centroids helps the algorithm to converge to the final centroid faster than the non-iterative version. Fig. 4 shows the iterative KM algorithm.

For the KM algorithm, the procedure to find the left and right centroid c_l and c_r is the following:

(1) Sort all the discourse universe values y_i in ascending order, where $i = 1, 2, \cdots, N$, such $y_1 \le y_2 \le \cdots \le y_N$. Associate each y_i with its corresponding $\underline{\mu}_{\tilde{B}}(y_i)$ and $\bar{\mu}_{\tilde{B}}(y_i)$.

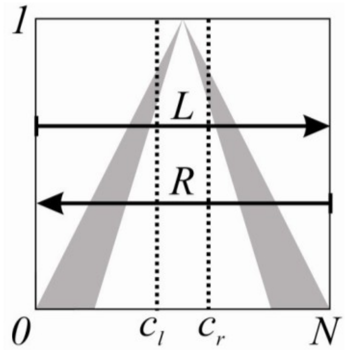


Fig. 3 The NI-KM algorithm performs all the centroid calculation.

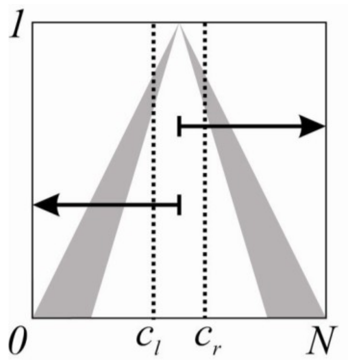


Fig. 4 The I-KM (iterative KM)algorithm. This algorithm is optimized for starting the centroid search from almost the half of discourse universe.

(2) Initialize

$$\theta_i = \frac{\underline{\mu}_{\tilde{B}}(y_i) + \overline{\mu}_{\tilde{B}}(y_i)}{2} \tag{5}$$

and compute

$$y = \frac{\sum_{i=1}^{N} y_i \theta_i}{\sum_{i=1}^{N} \theta_i} \tag{6}$$

(3) Find the switch point k, such $1 \le k \le N - 1$ and $y_k \le y_{k+1}$,

(4) Establish

For the left centroid:

$$\theta_r = \begin{cases} \overline{\mu}_{\tilde{B}}(y_i) & i \le k \\ \underline{\mu}_{\tilde{B}}(y_i) & \text{otherwise} \end{cases}$$

For the right centroid:

$$\theta_i = \begin{cases} \underline{\mu}_{\tilde{B}}(y_i) & i \le k \\ \overline{\mu}_{\tilde{B}}(y_i) & \text{otherwise} \end{cases}$$

and compute

$$y = \frac{\sum_{i=1}^{N} y_i \theta_i}{\sum_{i=1}^{N} \theta_i}$$

(5) If $y' = y$. If true, then stop and assign $c_l = y$ or $c_r = y$, correspondingly, else continue.

(6) Assign $y = y$, y go to step 3.

In both versions, the generalized centroid is calculated simply by averaging both c_l and c_r centroids:

$$y = \frac{c_l + c_r}{2} \tag{7}$$

3.3 Enhanced KM Algorithm

To improve the KM algorithm calculations, the E-KM (enhanced KM) algorithm [3] does not start only at the initial embedded set; it also starts from a convenient switch point (Fig. 5).

The algorithm to find the left centroid c_l is the following:

(1) Sort all the discourse universe values x_i in ascending order, where $i = 1, 2, \cdots, N$, such $y_1 \le y_2 \le \cdots \le y_N$. Associate each y_i with its corresponding $\underline{\mu}_{\tilde{B}}(y_i)$ and $\overline{\mu}_{\tilde{B}}(y_i)$.

(2) Establish k = round (N/2.4) and compute

$$a = \sum_{i=1}^{k} y_i \overline{\mu}_{\tilde{B}}(y_i) + \sum_{i=k+1}^{N} y_i \underline{\mu}_{\tilde{B}}(y_i)$$

$$b = \sum_{i=1}^{k} \overline{\mu}_{\tilde{B}}(y_i) + \sum_{i=k+1}^{N} \underline{\mu}_{\tilde{B}}(y_i)$$

$$y = \frac{a}{b}$$

(3) Find the switch point $k' \in [1, N\text{-}1]$, such

$$y_{k'} \le y \le y_{k'+1} \tag{8}$$

(4) Check if $k' = k$. If true, stop and assign $c_l = y$; else continue.

(5) Compute

$$a' = a \pm \sum_{i=\min(k,k')}^{\max(k,k')} y[\overline{\mu}_{\tilde{B}}(y_i) - \underline{\mu}_{\tilde{B}}(y_i)]$$

$$b' = b \pm \sum_{i=\min(k,k')+1}^{\max(k,k')} [\overline{\mu}_{\tilde{B}}(y_i) - \underline{\mu}_{\tilde{B}}(y_i)]$$

and compute again

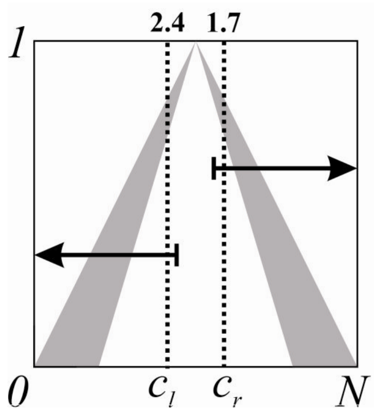


Fig. 5 The E-KM algorithm. This algorithm is doubly optimized for starting the centroid search eight percent around N/2 of discourse universe. That is why some literature uses the values 1.7 and 2.4 to start the centroid search.

$$y' = \frac{a'}{b'}$$

(6) Assign $y = y'$, $a = a'$, $b = b'$, $k = k'$ and go to step 3.

Each preliminary definition of $k_l = N/2.4$ and $k_r = N/1.7$ suggests that the initial centroids can be found ±8% around the middle of the set support, i.e., $k_l = (N/2) - 0.08N$ and $k_r = (N/2) + 0.08N$. These values were obtained experimentally [3]. This initial search reduces the search of the final centroid.

3.4 Nie-Tan Method

Although the NT method is not a KM algorithm derivation, this is used in the iterative-version KM algorithms to find the initial switching point. This method is discussed in this work because of its simplicity. The simplicity of the algorithm is based on the average between the LMF and UMF as it is shown in Fig. 6.

The NT method [9, 10, 11] is the simplest and fastest defuzzification method for IT2FLS.

It searches the middle MF θ_i between UMF and LMF. Its centroid is the approximated generalized centroid:

$$\theta_i = \frac{\overline{\mu}_{\tilde{B}}(y_i) + \underline{\mu}_{\tilde{B}}(y_i)}{2} \tag{9}$$

$$y = \frac{\sum_{i=1}^{N} y_i \cdot \mu_\theta(y_i)}{\sum_{i=1}^{N} \mu_\theta(y_i)} \tag{10}$$

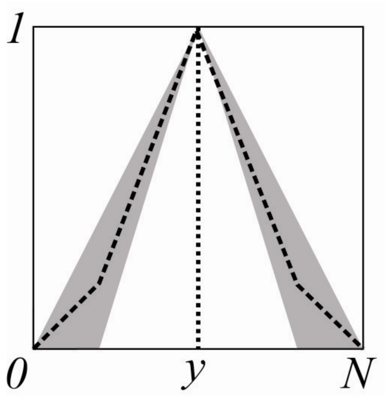


Fig.6 The NT method.

This method is one of the fastest methods because it requires only a division by 2 (1-position right shift register) and only a single centroid calculation.

Following section explains a hardware overview in FPGA in which the KM algorithms and the NT method were implemented.

4. DC Servo-Motor

This paper presents a hardware study between several defuzzification methods for IT2FLS. For doing this, the DC Servo-Motor application is proposed. The following section provides some theories about the DC servo-motor plant which is controlled by the IT2FLS.

The DC servo-motor model is well defined by conventional differential equations according to the Fig. 7. This model contains the complete description of the servomotor. Normally, in the control of DC servo-motor appears noise from the sensors, so a controller that could get rid of the noise effect is able to improve the whole performance of the system. The following figure shows the basic topology of the DC motor model.

The equations that describe the DC motor performance are presented below:

$$T = K_e \cdot i \tag{11}$$

$$e_a = K_f \cdot w_m \tag{12}$$

$$T = J \cdot \frac{\mathrm{d}}{\mathrm{d}t} w_m + b \cdot w_m \tag{13}$$

$$L \cdot \frac{\mathrm{d}}{\mathrm{d}t} i + R \cdot i = V - e_a \tag{14}$$

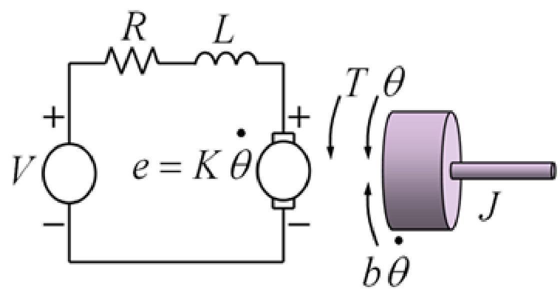


Fig. 7 DC servo-motor model.

where,

T: motor torque;
K_e: torque constant;
K_f: voltage constant;
i: armature current;
e_a: generated voltage;
w_m: angular velocity;
J: inertia;
b: damping ratio;
L: armature inductance;
R: armature resistance;
V: input voltage.

4.1 Laplace Transform Model

Rearranging Eqs. (11)-(14) :

$$K \cdot i(t) - b \cdot w_m(t) = J \cdot \frac{\mathrm{d}}{\mathrm{d}t} w_m(t) \tag{15}$$

$$V(t) - K \cdot w_m(t) - R \cdot i(t) = L \cdot \frac{\mathrm{d}}{\mathrm{d}t} i(t) \tag{16}$$

Our principal goal is to control the DC servo-motor speed by changing the input voltage V, so the Eqs. (15) and (16) can be transformed in terms of voltage V and angular speed w_m.

$$i(t) = \frac{J}{K} \cdot \frac{\mathrm{d}}{\mathrm{d}t} w_m(t) + \frac{b}{K} \cdot w_m(t) \tag{17}$$

Replacing Eq. (17) in Eq. (16) we get,

$$\left(\frac{L \cdot J}{K}\right) \cdot \frac{\mathrm{d}^2}{\mathrm{d}t^2} w_m(t) + \left(\frac{L \cdot b + R \cdot J}{K}\right) \cdot \frac{\mathrm{d}}{\mathrm{d}t} w_m(t) + \left(\frac{K^2 + R \cdot b}{K}\right) \cdot w_m(t) = V(t) \tag{18}$$

Before applying Laplace transform is necessary to represent the last equation in terms of deviation variables,

$$\left(\frac{L \cdot J}{K}\right) \cdot \frac{\mathrm{d}^2}{\mathrm{d}t^2} W_m(t) + \left(\frac{L \cdot b + R \cdot J}{K}\right) \cdot \frac{\mathrm{d}}{\mathrm{d}t} W_m(t) + \left(\frac{K^2 + R \cdot b}{K}\right) \cdot W_m(t) = \forall(t) \tag{19}$$

where,

$$W_m(t) = w_m(t) - w_m(0)$$

$$\forall(t) = V(t) - V(0)$$

And $w_m(0)$ and $V(0)$ are the initial conditions.

Applying the Laplace transform we get,

$$\left(\frac{L \cdot J}{K}\right) \cdot s^2 W_m(s) + \left(\frac{L \cdot b + R \cdot J}{K}\right) \cdot s W_m(s) + \left(\frac{K^2 + R \cdot b}{K}\right) \cdot W_m(s) = \forall(s) \tag{20}$$

Rearranging Eq.(20) we obtain the transfer function that represents the servomotor plant in terms of Laplace transform as follows:

$$\frac{W_m(t)}{\forall(s)} = \frac{1}{\left(\frac{L \cdot J}{K}\right) \cdot s^2 + \left(\frac{L \cdot b + R \cdot J}{K}\right) \cdot s + \left(\frac{K^2 + R \cdot b}{K}\right)} \tag{21}$$

4.2 State-Space Ttransfer Function

From Eq. (18), it can be defined the state-space model introducing the next variables,

$$x_1 = w_m(t) \tag{22}$$

$$x_2 = \frac{\mathrm{d}}{\mathrm{d}t} w_m(t) \tag{23}$$

The model is defined as,

$$\dot{x}_1 = \frac{\mathrm{d}}{\mathrm{d}t} w_m(t) = x_2 \tag{24}$$

$$\dot{x}_2 = \frac{\mathrm{d}^2}{\mathrm{d}t^2} w_m(t) = \left(\frac{K}{L \cdot J}\right) \cdot V(t) - \left(\frac{L \cdot b + R \cdot J}{L \cdot J}\right) \cdot x_2 - \left(\frac{K^2 + R \cdot b}{L \cdot J}\right) \cdot x_1 \tag{25}$$

$$y = w_m(t) = x_1 \tag{26}$$

$$u = V(t) \tag{27}$$

Finally, the state-space equations are,

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\left(\frac{K^2 + R\cdot b}{L\cdot J}\right) & -\left(\frac{L\cdot b + R\cdot J}{L\cdot J}\right) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{K}{L\cdot J}\right) \end{bmatrix} [u] \tag{28}$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{29}$$

And the DC motor model constants are,

$$A = \begin{bmatrix} 0 & 1 \\ -\left(\frac{K^2 + R\cdot b}{L\cdot J}\right) & -\left(\frac{L\cdot b + R\cdot J}{L\cdot J}\right) \end{bmatrix} \tag{30}$$

$$B = \begin{bmatrix} 0 \\ \left(\frac{K}{L\cdot J}\right) \end{bmatrix} \tag{31}$$

$$C = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{32}$$

$$D = 0 \tag{33}$$

4.3 Servo-Motor Control System

The servo-motor control system for the defuzzification method performance comparison is shown in Fig. 8, where the IT2FLS performs the position control using the position error θ and the change of the position θ; also, the servomotor plant can be represented by Eq. (21) or Eqs. (30) and (33). Fig. 9presents some details about the IT2FLS and its rules; nine rules are performed relating the antecedents to infer the consequences.

5. The Hardware Complexity

A real-time application can be managed by a computational device but the device selection will depend on the latency of that digital system that will manage it. A real-time application requires fast digital systems (with low latency) because its real-time characteristic will be defined by its WCET (worst case execution time).

The centroid calculation had been considered as a problem in T1FLS, for real-time applications because of its large latency. Although several applications were solved without problems with the centroid calculation in T1FLS, the use of this in the KM algorithms are very common and intensive and therefore the IT2FLS are hugely more complex than the T1FLS. The search of each centroid for each embedded set is globally several times more complex compared with the T1FLS.

This is one of the reasons why the IT2FLS are not used as the T1FLS today, and its robustness capability is not approached. The IT2FLS are often implemented in hardware such as RT-MCU (real-time microcontroller units), or DSP (digital signal processors) [12] and FPGA [13], because the rest of the hardware (computers with sequential program execution) becomes impractical and not feasible for real-time applications. These digital systems are often more expensive, especially the FPGA, and commonly this economic reason limits the application implementation in most cases.

Some of the high-end FPGA hardware elements that can be found in real-time applications are:

- high-speed dedicated multiplications;
- high-speed dedicated memory;
- DSP blocks;
- real-parallelism;
- partial reconfiguration.

Full-customizable multiplications are available in the FPGAs as a solution for improving the timing performance. Xilinx is the main FPGA brand that is used widely in several applications along the world. Some Xilinx FPGA families provide faster multiplication blocks and additional high performance resources (like DSP blocks and RT processors in the same land) in their higher end devices. Some high-speed RAM (random access memory) blocks are available for storing data, for instance.

These resources help the expert to design high speed and low cost T1FLS and IT2FLS in hardware among the software approaches, because a sequential program is executed in a fixed computing architecture [5]. If the specific application deserves the use of IT2FLS, then the expert must take into account how to select the appropriated defuzzification method that will be used for its real-time application.

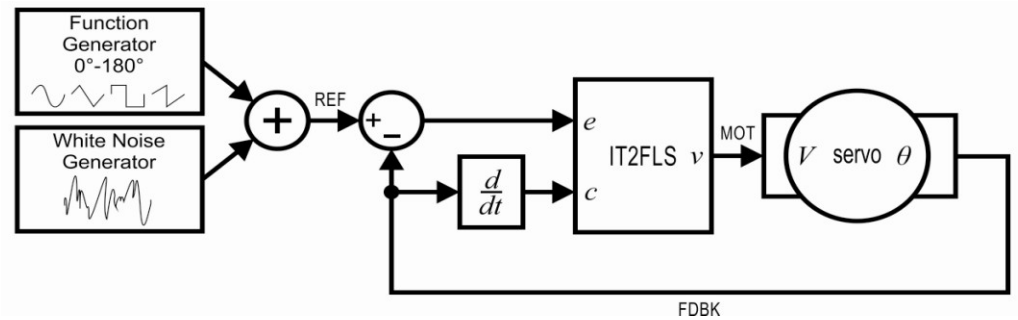


Fig. 8 The servo-motor control system.

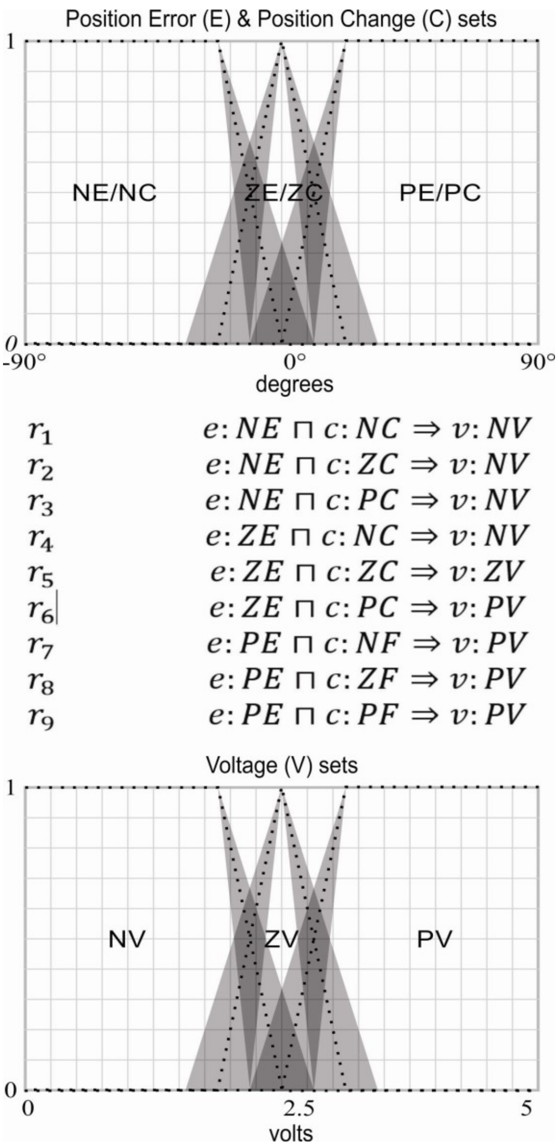


Fig. 9 The fuzzy sets and the rule set for the servo-motor application.

6. Methodology

The hardware defuzzification method comparisons are performed implementing all the KM algorithms, the NT method in FPGA hardware; National Instruments Reconfigurable I/O (compact RIO or cRIO) device is used and LabVIEW FPGA module to program every defuzzification stage enunciated in this paper and all the IT2FLS architecture. The cRIO device used is the NI-9014 with analog I/O capabilities, the C-series NI-9263 and the NI-9201 modules.

For this reason, the expert may also know some hardware details about the KM algorithms that will use.

For hardware comparison, the DC servo-motor application is proposed as a study case where the rotor position θ is tracked when a signal generator is introduced and a noise generator is used to disturb the IT2FLS performance, according to Fig. 8. Although the servo case is relatively slow, all these defuzzification methods can be analyzed and compared. Several studies have been implemented in FPGA hardware for T1FLS [14] and IT2FLS [15]. However, no comparisons have been presented, in hardware terms, where the most important defuzzification methods are applied for solving the same problem.

The DC servo-motor application, as can be seen in Figs. 8 and 10, is an electronic control training module dedicated for the cRIO device where the user can test control performance easily. The servomotor moves the rotor position from 0° to 180° according to the REF analog signal; the REF signal is used in the cRIO as an input to determine the reference that the controller will follow. The FDBCK (feedback) signal is another

analog signal used in the cRIO as an input also; the difference is that this signal is used to know where the current rotor position is.

The control system described in Fig. 8 is implemented using the LabVIEW FPGA module, where several diagram blocks are used for the data acquisition (input nodes), the noise generator summed to the reference signal (signal generator), the data driving to the servo (output node), and finally the IT2FLS. The error is computed as the difference between the desired values, such the REF (reference) signal, i.e., the reference and the current position value FDBCK, as can be seen in Figs. 8 and 10.

The cRIO analog output is used to drive the desired voltage to the servo-motor and this way determine the new position of the rotor. This signal is applied in the MOT (motor) signal as can be seen in the training module of the Fig. 10.

Details about the servo-motor control training module are not known, although their general equations are defined (From Eqs.(20) to (29)). This is not important for the fuzzy controller, because the IT2FLS must determine its control law without knowing the exact system dynamics.

For comparison purposes, the noisy reference signal REF is used so that the IT2FLS can follow it, the fuzzification and inference processes are previously tuned and are the same for the four cases. Four signal shapes were tested with the four methods: the NI-KM algorithm, the I-KM algorithm, the E-KM algorithm and the NT method, as can be seen in Fig. 11.

7. Results and Discussions

7.1 Reference Tracking

The final response of the control system using all the defuzzification methods mentioned is shown in Fig. 11. Four signal shapes were used to the IT2FLS tracking.

The NI-KM algorithm is the worst case and its latency makes the decision to take late decisions. The best results were obtained with the E-KM algorithm and the NT method. This behavior is similar in each case.

7.2 The Hardware Performance

The following paragraphs are dedicated to describe the hardware complexity between every defuzzification method described in previous sections. NI-cRIO devices are used in conjunction with NI LabVIEW FPGA module.

The following section provides a timing and area performance analysis where each KM algorithm is analyzed and compared. Also, the NT method is included.

7.2.1 Complexity and Arithmetic-related

Assume that each algorithm requires two memories with N locations and each centroid (either left or right) are computed in parallel. Each element is a byte (8-bit width). Every memory can be distributed in the FPGA by LUT (look-up tables).

The following chart provides a comparison between all the KM algorithms and the NT method.

According to Eqs.(1-9), the total iteration count can be obtained as an approximation if a counter is included in the most inner loop of the algorithm and by observation of the block diagram. In this case, N = 256, where N is the number of points the discourse universe is divided into. So, the non-iterative KM may

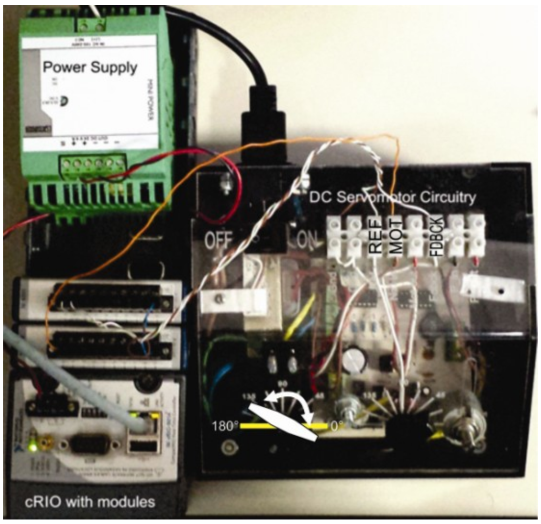


Fig. 10 The control of a servo-motor training module and the cRIO with C-series modules.

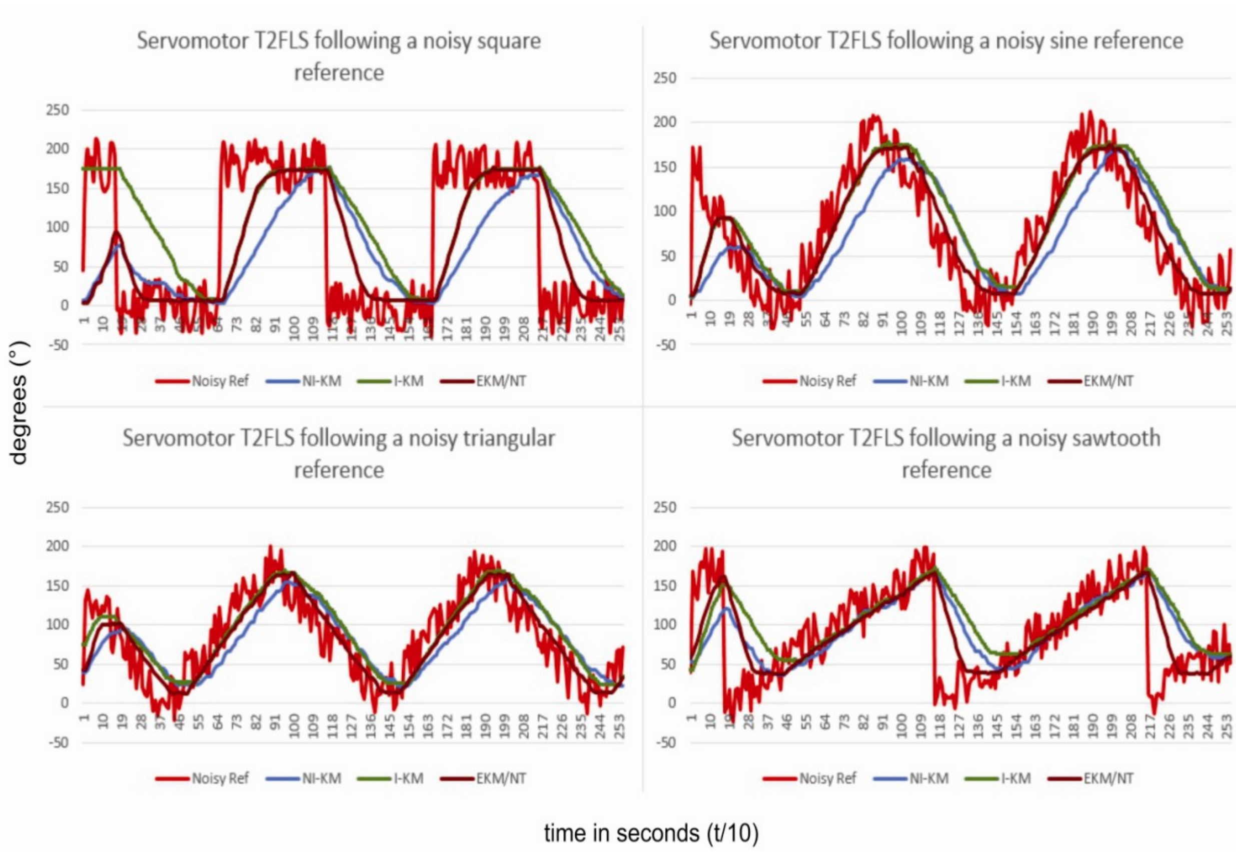


Fig. 11 All the KM algorithms and NT method tracking performance comparison in presence of noise. Four signal shapes were used to be tracked by the IT2FLS using each algorithm. The total time for each experimental test is 25.5 seconds.

last $2N^2 + 4N = 2(256)^2 + 4(256) = 132.096$ iterations for a single defuzzification without taking into account the iterations required for the T2 fuzzification and T2 inference processes, as can be seen in Table 1. For instance, the E-KM requires from 1 up to $0.1764N^2 + 2.1N = 0.1764(256)^2 + 2.1(256) = 12.098$ iterations to find the final left or right centroid. Due to each left or right centroid calculation is processes in parallel; it can be seen as a single loop. The iterative version, which can be compared with the non-iterative version and may last the same time to execute with the difference of the early termination condition, converges slower than the E-KM.

The NT method also finishes before, because its implementation requires only a single centroid calculation and only a single embedded set in the set FOU. Although the NT method seems to be the fastest, the E-KM may find the final centroid in the first iteration, which very significant compared with the NT algorithm.

Table 1 summarizes the complexity of each defuzzification method, which was obtained experimentally.

7.2.2 Resource Usage

The cRIO device provides several limited resources like multiplications, memories, amongst others which provide very high performance and the expert may use to build more complex structures like divisions or square roots. LabVIEW FPGA module is a GUI (graphical user interface) which provides several basic tools that can be used to implement high-throughput operations in a RIO device, controls that let the user modify the input data to the digital system and indicators that let the user show the output data from the digital system, related to a virtual instrument that works as the programming unit.

In functional block terms (LabVIEW FPGA module), the KM algorithms and the NT method were designed, implemented are compared according to the structures used for being implemented.

After the module compilation, LabVIEW FPGA provides the FPGA resulting resource utilization. As can be seen in the last chart (Table 2), the resource usage can be useful for comparison purposes. This way, can be seen that the E-KM is the most expensive, but this is the fastest defuzzification method, but the NT method is the cheapest method because requires only one division and one multiplication for calculating the final centroid. Because this method does not search for several embedded sets, the final centroid is a single value and that is why requires of a single centroid calculation unit. The E-KM and the I-KM require 3 centroid calculation units while the non-iterative version requires 2.

7.2.3 Timing and Area Resource Usage

Also, the final timing and resources of each method is presented using the LabVIEW Tick Count block. And the resource utilization is obtained from the Build Specifications in LabVIEW FPGA module.

The best timing performance is reached by the E-KM algorithm and the second best one is the NT, although the E-KM also presents the worst resource utilization. Now observe that the non-iterative KM algorithm presents the worst case, achieving about 20 FLIPS (fuzzy logic inferences per second), which is not practical for the IT2FLS applied in real-time questions. Also, the NT method presents the best resource utilization. The software VI timing performance depends on the operating system tick time, which is generally 55 milliseconds per tick, so for the NI-KM the total ticks are 1.537, then $1.537 \times 55ms = 84.535ms$, as can be seen in Table 3.

LabVIEW FPGA module is used as reference for testing hardware and performing comparisons. This module creates a whole computing architecture that generates great amount of slices for any design. Its advantage stands on the fact of the ease of hardware validation due to its graphical interface, characteristic of the virtual instruments in LabVIEW.

The final resource utilization may vary if the system is implemented directly in a Xilinx FPGA and without using the LabVIEW FPGA module.

Table 1 Number of iterations per defuzzification method.

Element/method	NI-KM	I-KM	E-KM	NT
ES calculation	N	N	N/A	N
ES centroid calculation	$N + 3$	$N + 3$	$N + 1$	$N + 3$
Total iteration count	$2N^2 + 4N$	$[N/2, 2N^2 + 4N]$	$[1, 0.1764N^2 + 2.1N]$	$2N + 3$

Table 2 Number of hardware elements used for each defuzzification method.

Structure/method	NI-KM	I-KM	E-KM	NT
Multiplications	1	1	6	1
Divisions	1	2	5	1
Sums/subtracts	3	4	17	5
Centroid calculation units	2	3	3	1
MUX (Comparator/multiplexers)	5	9	9	1

Table 3 Timing performance and resource utilization per defuzzification method.

Resource/methods	NI-KM	I-KM	E-KM	NT
Latency (hardware) in milliseconds	49.48	0.8875	0.1756	0.27
Latency (software) in milliseconds	84.535	1.87	1.43	1.32
Slices	1,461	2,415	2,593	915
Registers	1,454	2,087	2,828	959
LUT	2,185	3,759	3,965	1,305

8. Conclusions

The E-KM algorithm and the NT Method are the best choices for implementing real-time control applications in hardware. The NI-KM is not practical due to its poor response. Then, depending on the expert requirements, the expert must choose between speed and resources, so the E-KM algorithm provides the best speed although not the best resource usage. The NT method is fast and simple in the number of resource for real-time control applications.
In the specific case of DC motors is an excellent option to use a the E-KM in the defuzzification stage, so it could be very useful to implement it in different industrial applications like CNC (computer numerical control) machines that run in real time.

Acknowledgments

This work is sponsored by Instituto Tecnologico de Estudios Superiores de Monterrey and National Instruments.

References

[1] J.M. Mendel, Type-2 fuzzy sets and systems: An Overview, IEEE Computational Intelligence Magazine 2 (1) (2007) 20-29.
[2] N.N. Karnik, J.M. Mendel, Q. Liang,Type-2 fuzzy logic systems, IEEE Transactions on Fuzzy Systems 7 (1999) 643-658.
[3] H. Wu, J. M. Mendel, Enhanced Karnik-Mendel algorithms, IEEE Transactions on Fuzzy Systems 17 (2009) 923-934.
[4] M.Biglarbegian, W. Melek, J.Mendel, Robustness of interval type-2 fuzzy logic systems, in: IEEE NAFIPS Canada, 2010.
[5] N. Manaresi, R. Rovatti, E. Franchi, R. Guerrieri, G. Baccarani, Automatic synthesis of analog fuzzy controllers: A hardware and software approach, IEEE Transactions on Industrial Electronics 43 (1) (1996) 217-225.
[6] J. Mendel, X. Liu, Some extensions of the Karnik-Mendel algorithms for computing an interval type-2 fuzzy set centroid, in: IEEE Symposium on Advances in Type-2 Fuzzy Logic Systems, Paris, France, April 11-15, 2011.
[7] J. Mendel, F. Liu, Super-Exponential convergence of the Karnik-Mendel algorithms used for type-reduction in interval type-2 fuzzy logic systems, in: IEEE International Conference on Fuzzy Systems, Vancouver, BC, Canada, July 16-21, 2006.
[8] J. Mendel, R.I.B. John, Type-2 fuzzy sets made simple, IEEE Transactions on Fuzzy Systems 10 (2) (2002) 117-127.
[9] J.M. Mendel, X. Liu, Simplified interval type-2 fuzzy logic systems, IEEE Transactions on Fuzzy Systems 21 (6) (2013) 1056-1069.
[10] D. Wu, J. M. Mendel,On the Continuity of type-1 and interval type-2 fuzzy logic systems, IEEE Transactions on Fuzzy Systems 19 (1) (2011) 179-192.
[11] M. Nie, W.W. Tan, Towards an efficient type-reduction method for interval type-2 fuzzy logic systems, in: IEEE World Congress on Computational Intelligence, Hong Kong, China, June 1-6, 2008.
[12] R. Lauwereins, M. Engels, J.A. Peperstraete, Parallel processing enables the real-time emulation of DSP ASICs, IEEE International Workshop on Rapid System Prototyping, North Carolina, USA, June 4-7, 1990.
[13] Y. Chen, V. Dinavahi, Multi-FPGA digital hardware design for detailed large-scale real-time electromagnetic transient simulation of power systems, IET Generation, Transmission & Distribution 7 (5) (2013) 451-463.
[14] M. Cirstea, J. Khor, M. McCormick, FPGA fuzzy logic controller for variable speed generators, in: IEEE International Conference on Control Applications, Mexico, Sep. 5-7, 2001.
[15] M. Melgarejo, C.A. Pena-Reyes, Implementing interval type-2 fuzzy processors, IEEE Computational Intelligence Magazine 2 (1) (2007) 63-71.

Journal of Control Science and Engineering 1 (2013) 13-22

Extinction and Persistent of a Stochastic Multi-group SIR Epidemic Model

Xiaojing Zhong and Feiqi Deng

Systems Engineering Institute, South China University of Technology, Guangzhou 510640, P.R. China

Abstract: We establish a stochastic differential equation epidemic model of multi-group SIR type based on the deterministic multi-group SIR mode. Then, we define the basic reproduction number R_0^S and show that it is a sharp threshold for the dynamic of the stochastic multi-group SIR model. More specially, if $R_0^S < 1$, then the disease-free equilibrium will be asymptotically stable which means the disease will die out, if $R_0^S > 1$, the disease-free equilibrium will unstable, and our model will positively recurrence to a positive domain which implies the persistence of our model. Numerical simulation examples are carried out to substantiate the analytical results.

Key words: Stochastic, multi-group SIR model, threshold dynamics, positive recurrence, stochastic persistence.

1. Introduction

For the past decades, many epidemic models have been proposed for modeling the spread process of infectious diseases, and in the meantime considerable attention has been paid to study the dynamical properties of these various models. Most models descend from the classical SIR model of Kermack and McKendrick [1], and then different type of epidemic models have been researched by many scholars [2-10]. In particular, multi-group models have been proposed to describe the transmission dynamics of infectious diseases in heterogeneous host populations, such as meals, mumps, gonorrhea, HIV/AIDS, WNV (West-Nile virus) and vector borne diseases such as malaria. One of the earliest works on multi-group models is the seminal paper by Laj-manovich and Yorke [11] on a class of SIS (suspectible, infected, suspectable) multi-group models for the transmission dynamics of Gonorrhea, which established a complete analysis of the global dynamics. The global stability of the unique equilibrium is proved by using a complete analysis of the global Lyapunov function. Subsequently, much research has been done on multi-group models [12-18]. Recently, a group-theoretic approach to the method of global Lyapunov function was proposed by Michael et al. [19].In the paper, the authors studied the following SIR model:

$$\begin{cases} S_k' = A_k - d_k^s S_k - \sum_{j=1}^{n} \beta_{kj} S_k I_j, \\ I_k' = \sum_{j=1}^{n} \beta_{kj} S_k I_j - (d_k^I + \varepsilon_k + \gamma_k) I_k, \\ R_k' = \gamma_k I_k - d_k^R R_k, \end{cases} \quad (1)$$

The model describes the spread of an infectious disease in a heterogeneous population, which is partitioned into n homogeneous group. Each group k is further compartmentalized into S_k, I_k and R_k, here S_k, I_k and R_k denote the susceptible, infective and recovered population at time t, respectively. All parameters in the above model are nonnegative constants and summarized in the following list:

β_{ij}: transmission coefficient between compartments S_i and I_j;

Corresponding author: Feiqi Deng, Professor, research fields: stability, stabilization and robust and variable structure control theory of complex systems, including time-delay systems, nonlinear systems and stochastic systems, and machine learning.E-mail:aufqdeng@scut.edu.cn.

d_k^S, d_k^I, d_k^R :nature death rates of S, I, R compartments in the k-th group, respectively;

A_k: influx of individuals into the k-th group;

γ_i: recovery rate of infectious individuals in the i-th group;

ε_k: disease-caused death rate in the k-th group.

All parameter values are assumed to be nonnegative and $d_k^S, d_k^I, d_k^R, A_k > 0$ for all k. According to Michael et al. [19], there is a disease-free equilibrium: $P_0 = (S_1^0, 0, 0, \cdots, S_n^0, 0, 0)$ on the boundary of Γ, where $S_i^0 = A_i / d_i^S$ and

$$\Gamma = \{(S_1, I_1, \cdots, S_n, I_n) \in R_+^{2n} \Big| S_k \le \frac{A_k}{d_k^S},$$
$$S_k + I_k \le \frac{A_k}{\min\{d_k^S, d_k^I + \varepsilon_k, d_k^R\}}\} \quad (2)$$

A threshold R_0 is defined which decide the epidemic will prevalent or not, where

$$R_0 = \rho(M_0) \quad (3)$$

denote the spectral radius of the matrix

$$M_0 = M(S_1^0, S_2^0, \ldots, S_n^0) = \left(\frac{\beta_{ij} S_i^0}{d_i^I + \gamma_i + \varepsilon_i}\right), 1 \le i, j \le n \quad (4)$$

Here we give more details. If $R_0 \le 1$, then, P_0 is the unique equilibrium and it is globally asymptotically stable in Γ; If $R_0 > 1$, then P_0 is unstable and it is uniformly persistent. Furthermore, there exists an endemic equilibrium P^* and it is globally asymptotically stable in Γ. In the whole proof, a very important graph theorem was used.

Given a nonnegative matrix $A = (\beta_{ij})$ the directed graph G(A) associated with $A = (\beta_{ij})$ has vertices 1, 2, ···, n with a directed arc (i, j) from i to j if $\beta_{ij} \ne 0$. It is strongly connected if any two distinct vertices are joined by an oriented path. The matrix A is irreducible if and only if G(A) is strongly connected. A tree is a connected graph with no cycles. A sub tree T of a graph G is said to be spanning if T contains all the vertices of G. A directed tree is a tree in which each edge has been replaced by an arc directed one way or the other. A directed tree is said to be rooted at a vertex, called the root, if every arc is oriented in the direction towards the root. An oriented cycle in a directed graph is a simple closed oriented path. A unicyclic graph is a directed graph consisting of a collection of disjoint rooted directed trees whose root are on an oriented cycle. For a given nonnegative matrix A=(a_{ij}), let:

$$L = \begin{bmatrix} \sum_{l\ne 1}\beta_{1l} & -\beta_{21} & \cdots & -\beta_{n1} \\ -\beta_{12} & \sum_{l\ne 2}\beta_{2l} & \cdots & \beta_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ -\beta_{1n} & -\beta_{2n} & \cdots & \sum_{l\ne n}\beta_{nl} \end{bmatrix} \quad (5)$$

be the Laplacian matrix of the directed graph G(A) and C_{ij} denote the cofactor of the (i, j) entry of L. In light of these results, complete determination of the global dynamic of these models is essential for their application and further development.

Whereas the statement above, the large-scale biological system's parameters are assumed as constants, but in the real situation, parameters involved with the model always fluctuate around some average value due to the continuous fluctuation in the environment. In order to study the dynamics of interacting population under realistic situation, we need to analyse the associated stochastic model. In fact, the stochastic epidemic models have been studied by many authors [20-36], they established related stochastic epidemic model based on the deterministic model. By using the Lyanupov method, Tornatore et al.[30],Yu et al. [31], Ji et al. [32], Liu et al. [33] and Ji et al. [34] found out sufficient conditions of the stability of the steady-state based on the deterministic threshold R_0. Gray et al. [35] established a stochastic SIS model and found out the sufficient and necessary condition of the disease-free equilibrium and the condition of the persistence of the disease. Liu et al.[37] gave many stochastic persistence definitions about epidemic model.

In this paper, we perturb the death rate and transmission coefficient of the deterministic multi-group SIR model Eq. (1) by replacing d_k^I, d_k^I and β_{kk}, by$d_k^I + \sigma_I \dot{B}(t), d_k^R + \sigma_R \dot{B}(t)$ and $\beta_{kk} + \sigma_k \dot{B}(t)$

where, $B_I(t), B_R(t), B_\beta(t)$ are independent standard Brownian motions with $B_I(0) = 0, B_R(0) = 0$, $B_\beta(0) = 0$. Then we formulate a stochastic multi-group SIR model as follows:

$$\begin{cases} \mathrm{d}S_k = (A_k - d_k^S S_k - \sum_{j=1}^{n} \beta_{kj} S_k I_j)\,\mathrm{d}t - \sigma_k S_k I_k\,\mathrm{d}B_\beta, \\ \mathrm{d}I_k = [\sum_{j=1}^{n} \beta_{kj} S_k I_j - (d_k^I + \varepsilon_k + \gamma_k) I_k]\,\mathrm{d}t \\ \quad + \beta_{kk} S_k I_k dB_\beta - \sigma_I I_k\,\mathrm{d}B_I, \\ \mathrm{d}R_k = (\gamma_k I_k - d_k^R R_k)\,\mathrm{d}t - \sigma_R R_K\,\mathrm{d}B_R. \end{cases} \tag{6}$$

We prove the global existence of the positive solution in Section 2. The stability of the disease-free equilibrium is derived in Section 3, we will find out a new threshold R_0^S different from the deterministic R_0 which determine the extinction and persistence of the disease. In Section 4, we give the proof of our model's stochastic persistence. Examples and the simulation results are considered to illustrate our main results.

Throughout the article, unless otherwise specified, we will employ the following notions. Let $(\Omega, F, \{F_t\}_{t\geqslant 0}, P)$ be a complete probability space with a filtration $\{F_t\}_{t\geqslant 0}$ satisfying the usual conditions, i.e., it is right continuous and F_0 contains all P-null sets. We use $a \vee b$ to denote max(a, b), $a \wedge b$ to denote min(a, b) and a.s. to mean almost surely.

2. Existence and Uniqueness of the Positive Global Solution

In this section, we prove the global existence of the positive solution of our stochastic system Eq. (6). As a stochastic differential equation, the functions involved with stochastic system are generally required to satisfy the Lipschitz condition and linear growth condition. Obviously, the functions in Eq. (6) do not satisfy the linear growth condition, so the solutions may explode at finite times. To solve this problem, we first show that the local positive solutions exist before the explode times, then we use the lyapunov function method to prove that the solutions exist globally.

Theorem 2.1 If $B=(\beta_{ij})_{n\times n}$ is irreducible, then for any initial value $(S_1(0), I_1(0), R_1(0), \cdots, S_n(0), I_n(0), R_n(0)) \in R_+^{3n}$, there exists a unique solution $(S_1(t), I_1(t), R_1(t), \cdots, S_n(t), I_n(t), R_n(t)) \in R_+^{3n}$ to system Eq. (6) and it satisfies $P((S_1(t), I_1(t), R_1(t), \cdots, S_n(t), I_n(t), R_n(t) \mid (S_1(t), I_1(t), R_1(t), \cdots, S_n(t), I_n(t), R_n(t) \in R_+^{3n})$ =1, Which mean $(S_1(t), I_1(t), R_1(t), \cdots, S_n(t), I_n(t), R_n(t) \in R_+^{3n}$.

Proof. Since the coefficients of the equation are locally Lipschitz continuous, there is unique local solution on $t \in [0, \tau_e)$, where τ_e is the explosion time [38].Using the Itô formula, the solution can be expresses as:

$$S_k(t) = e^{-d_k^S t - \int_0^t (\sum_{j=1}^{n} \beta_{kj} I_j(u) + \frac{\sigma_k^2}{2} I_k^2(u))\mathrm{d}u - \sigma_k \int_0^t I_k(u)\mathrm{d}B_\beta(u)}$$
$$\times [S_k(0) + A_k \int_0^t e^{d_k^S u + \int_0^u (\sum_{j=1}^{n} \beta_{kj} I_j(v) + \frac{\sigma_k^2}{2} I_k^2(v))\mathrm{d}v + \sigma_k \int_0^u I_k(v)\mathrm{d}B_\beta(v)}\,\mathrm{d}u] \tag{7}$$

then $S_k(t) > 0, t \in [0, \tau_e)$, By the same way we get

$$I_k(t) = e^{-(d_k^I + \varepsilon_k + \gamma_k + \frac{\sigma_I^2}{2})t + \int_0^t (\beta_{kk} S_k(u) - \frac{\sigma_k^2}{2} S_k^2(u))\mathrm{d}u}$$
$$\times e^{\sigma_k \int_0^t S_k(u)\mathrm{d}B_\beta(\upsilon) + \sigma_I B_I(t)} \times \{I_k(0)$$
$$+ \int_0^t \sum_{k\neq j} \beta_{kj} S_k I_j(v) [e^{(d_k^I + \varepsilon_k + \gamma_k + \frac{\sigma_I^2}{2})u}$$
$$\times e^{-\int_0^u ((\beta_{kk} S_k(v) - \frac{\sigma_k^2}{2} S_k^2(v))\mathrm{d}v - \sigma_k \int_0^u S_k(v)\mathrm{d}B_\beta(v) + \sigma_I B_I(u)}]du\} \tag{8}$$

$$R_k(t) = e^{-(d_k^R + \frac{\sigma_R^2}{2})t - \sigma_R B_R(t)}$$
$$\times [R_k(0) + \int_0^t \gamma_k I_k(u) e^{(d_k^R + \frac{\sigma_R^2}{2})u - \sigma_R B_R(u)}\,\mathrm{d}u] \tag{9}$$

then we can conclude that $S_k(t)$, $I_k(t)$, $R_k(t)$ are positive on $t \in [0, \tau_e)$. Then we want to show that this solution is a global solution. To prove this, we need to show that $\tau_e = \infty$ almost surely. We choose a sufficiently large number 0 such that $S(0)$, $I(0)$, $R(0)$ all lie with the interval $(0, m_0)$, For each integer $m > m_0$, we define the stopping time:

$$T_m = \inf\{t \in [0, \tau_e) : \max\{S_k(t) + I_k(t) + R_k(t),\ k = 1,2,\cdots,n > m\} \tag{10}$$

where, $\inf\emptyset=\infty$. Set $\tau_\infty = \lim_{k\to\infty}$ whence $\tau_\infty < \tau_e$. If we can show that $\tau_\infty = \infty$ a.s. it is sufficient to prove that $\tau_e = \infty$ a.s. for all $t \geq 0$. If this statement were false, then there is a pair of constants $T > 0$ and $\varepsilon \in (0, 1)$ such that $P\{\tau_\infty \leq T\} > \varepsilon$, hence there is an integer $m_1 \geq m_0$ such that $P\{\tau_m \leq T\} \geq \varepsilon$ for all $m \geq m_1$.

Define a function $V(S_1,I_1,R_1,\cdots, S_n, I_n, R_n) = \sum_{k=1}^{n} (S_k+I_k+R_k)$,using the Itô formula, for any $t\in (0, T]$ and $m \geq m_1$,

$$\begin{aligned} &EV(S_1(t\wedge\tau_m), I_1(t\wedge\tau_m), R_1(t\wedge\tau_m),\cdots,S_n(t\wedge\tau_m),\\ &I_n(t\wedge\tau_m), R_n(t\wedge\tau_m)) = V(S_1(0), I_1(0), R_1(0),\cdots,\\ &S_n(0), I_n(0), R_n(0)) + E\int_0^{t\wedge\tau_m} LV(S_1(s), I_1(s), R_1(s),\cdots,\\ &S_n(s), I_n(s), R_n(s))ds, \end{aligned} \tag{11}$$

where, LV is

$$\begin{aligned} LV &= \sum_{k=1}^{n}[A_k - d_k^S S_k - (d_k^I + \varepsilon_k)I_k - d_k^R R_k]\\ &\leq \sum_{k=1}^{n} A_k = A \end{aligned} \tag{12}$$

Therefore, if $t \leq T$, we have

$$\begin{aligned} &EV(S_1(t\wedge\tau_m), I_1(t\wedge\tau_m), R_1(t\wedge\tau_m),\ldots,S_n(t\wedge\tau_m),\\ &I_n(t\wedge\tau_m), R_n(t\wedge\tau_m)) = V(S_1(0), I_1(0), R_1(0),\cdots,\\ &S_n(0), I_n(0), R_n(0)) + AT \end{aligned} \tag{13}$$

Set $\Omega_m = \{\tau_m \leq T\}$, for $m \geq m_1$, then we know $P(\Omega_m)$. For every $\omega \in \Omega_m$, $\max\{S_k(t) + I_k(t) + R_k(t),\ k = 1, 2,\cdots, n\} \geq m$, hence

$$\begin{aligned} &V(S_1(0), I_1(0), R_1(0),\ldots,S_n(0), I_n(0), R_n(0)) + AT\\ &\geq E[I_{\Omega_m}(\omega)V(S_1(t\wedge\tau_m), I_1(t\wedge\tau_m), R_1(t\wedge\tau_m),\cdots,\\ &S_n(t\wedge\tau_m), I_n(t\wedge\tau_m), R_n(t\wedge\tau_m))]\\ &\geq \varepsilon max\{S_k(t) + I_k(t) + R_k(t), k = 1,2,\cdots,n\}\\ &\geq \varepsilon m \end{aligned} \tag{14}$$

Letting $m\to\infty$ leads to the contradiction $\infty > V(S(0), I(0), R(0)) + AT > \infty$, so we have $\tau_\infty=\infty$ a.s. whence the proof is complete.

3. Extinction of the Epidemic

In the study of population systems, extinction and persistence are two of the most important issues. In this section, we will discuss the extinction of Eq. (6). Since the coexisting disease-free equilibrium P_0 of the deterministic SIR model Eq. (1), we make the variable changes $u_i(t) = S_k(t) - S_k^0$, $v_i(t) = I_k(t)$ and $w(t) = R_k(t)$, so that the origin will represent the disease-free equilibrium, by this we consider the linearized system:

$$\begin{cases} \mathrm{d}u_k = (-d_K^S u_k - \sum_{j=1}^{n} \beta_{kj}S_k^0 v_j)dt - \sigma_k S_k^0 v_k \,\mathrm{d}B_\beta(t),\\ \mathrm{d}v_k = [\sum_{j=1}^{n} \beta_{kj}S_k^0 v_j - (d_k^I + \varepsilon_k + \gamma_k)v_k]\mathrm{d}t\\ \quad - \sigma_I v_k \,\mathrm{d}B_I + \sigma_k S_k^0 v_k \,\mathrm{d}B_\beta(t),\\ \mathrm{d}w_k = [\gamma_k v_k - d_k^R w_i]\mathrm{d}t - \sigma_R w_k \,\mathrm{d}B_R(t). \end{cases} \tag{15}$$

Considering the second Eq. (15), let $x(t) = (v_1(t), v_2(t), \cdots, v_n(t))$, we rewrite the second equation as

$$\mathrm{d}x(t) = Fx(t)\mathrm{d}t + G_1 x(t)\mathrm{d}B_\beta(t) - G_2 x(t)\mathrm{d}B_I(t) \tag{16}$$

Where,

$$F = \begin{bmatrix} \beta_{11}S_1^0 - d_1^I - \varepsilon_1 - \gamma_1 & \cdots & \beta_{1n}S_1^0\\ \vdots & \ddots & \vdots\\ \beta_{nn}S_n^0 & \cdots & \beta_{n1}S_n^0 - d_n^1 - \varepsilon_n - \gamma_n \end{bmatrix} \tag{17}$$

$$G_1 = \begin{bmatrix} \sigma_1 S_1^0 & & & \\ & \sigma_2 S_2^0 & & \\ & & \ddots & \\ & & & \sigma_n S_n^0 \end{bmatrix} \tag{18}$$

$$G_1 = \begin{bmatrix} \sigma_I & & & \\ & \sigma_I & & \\ & & \ddots & \\ & & & \sigma_I \end{bmatrix} \tag{19}$$

We assume that $\sigma_k S_k^0 = \sigma_\beta$, then the matrices F, G_1, G_2 commute, the explicit solution of the linearized system in Eq.(16) can be solved as

$$x(t) = x(0)\exp[(F - \frac{1}{2}G_1^2 - \frac{1}{2}G_2^2)t + G_1 B_\beta(t) - G_2 B_I(t) \tag{20}$$

where,

$$F - \frac{1}{2}G_1^2 - \frac{1}{2}G_2^2 =$$

$$\begin{bmatrix} \beta_{11}S_1^0 - d_1^I - \varepsilon_1 - \gamma_1 - \frac{\sigma_I^2}{2} - \frac{(S_1^0\sigma_1)^2}{2} & \cdots & \beta_{1n}S_1^0 \\ \vdots & \ddots & \vdots \\ \beta_{nn}S_n^0 & \cdots & \beta_{n1}S_n^0 - d_n^I - \varepsilon_n - \gamma_n - \frac{\sigma_I^2}{2} - \frac{(S_n^0\sigma_n)^2}{2} \end{bmatrix} \tag{21}$$

Let $R_0^S = \rho(M_0^S)$ denote the spectral radius of the matrix

$$M_0^S = \left(\frac{\beta_{kj}S_k^0}{d_i^I + \varepsilon_k + \gamma_k + \frac{\sigma_k^2}{2} + \frac{(S_i^0\sigma_i)^2}{2}} \right)_{1 \le k, j \le n} \tag{22}$$

In this case, there is a pair of positive constants C and λ, so that

$$\left\| \exp[(F - \frac{1}{2}G_1^2 - \frac{1}{2}G_2^2)t] \right\| \le Ce^{-\lambda t} \tag{23}$$

It then follows from Eq.(23) that

$$|x(t)| \le C|x(0)| \exp[-\lambda t + \|G_I\| \|B_\beta(t)\| + \|G_2\| \|B_I(t)\|] \tag{24}$$

Using the strong law of large numbers, we get that $\lim_{t \to m} \frac{B_\beta(t)}{t} = 0$ a.s. and $\lim_{t \to m} \frac{B_I(t)}{t} = 0$ a.s., thus we obtain

$$\limsup_{t \to 0} \frac{1}{t} \log|x(t)| \le \lambda \quad \text{a.s.}$$

In other words, the solution of Eq.(15) is almost surely exponentially stable. Next, we give estimate for $u_k(t)$, $w_k(t)$. Using the Itˆo formula, we derive that

$$w_k(t) = e^{-(d_k^R + \frac{\sigma_R^2}{2})t - \sigma_R B_R(t)} [R_k(0) + \int_0^t \gamma_k v_k e^{-(d_k^R + \frac{\sigma_R^2}{2})u - \sigma_R B_R(u)} \mathrm{d}u] \tag{25}$$

Substituting Eq. (24) into Eq. (25) we get

$$w_k(t) \le e^{-(d_k^R + \frac{\sigma_R^2}{2})t - \sigma_R B_R(t)} [R_k(0) + \int_0^t \gamma_k v_k e^{-(d_k^R + \frac{\sigma_R^2}{2} - \lambda)u + \sigma_\beta |B_\beta(u)| + \sigma_I |B_I(u)| + \sigma_R B_R(u)} \mathrm{d}u] \tag{26}$$

By the law of the iterated logarithm,

$$\limsup_{t \to 0} \frac{B(s)}{\sqrt{2s \log\log s}} = 1 \quad \text{a.s.}$$

thus there is a $T > 0$ such that, if $t > T$, then we have $B(s) \le \sqrt{2s \log\log s}$. Using this for the estimation of $w_k(t)$, we obtain that

$$\begin{aligned} w_k(t) &\le e^{-(d_k^R + \frac{\sigma_R^2}{2})t - \sigma_R B_R(t)} [R_k(0) + \\ &\int_0^t \gamma_k v_k e^{(d_k^R + \frac{\sigma_R^2}{2} - \lambda)u + (\sigma_\beta \sigma_I + \sigma_R)\sqrt{2u \log\log u}} du] \\ &\le e^{-(d_k^R + \frac{\sigma_R^2}{2})t - \sigma_R B_R(t)} [R_k(0) + \\ &\int_0^T \gamma_k v_k e^{(d_k^R + \frac{\sigma_R^2}{2} - \lambda)u + (\sigma_\beta \sigma_I + \sigma_R)\sqrt{2u \log\log u}} du + \\ &+ \gamma_k e^{(\sigma_\beta \sigma_I + \sigma_R)\sqrt{2t \log\log t}} \int_T^t v_k e^{(d_k^R + \frac{\sigma_R^2}{2} - \lambda)u} du] \\ &= e^{-(d_k^R + \frac{\sigma_R^2}{2})t - \sigma_R B_R(t)} [R_k(0) + \\ &\int_0^T \gamma_k v_k e^{(d_k^R + \frac{\sigma_R^2}{2} - \lambda)u + (\sigma_\beta \sigma_I + \sigma_R)\sqrt{2u \log\log u}} du] \\ &+ \frac{\gamma_k}{d_k^R - \frac{\sigma_R^2}{2} - \lambda} e^{-\lambda t + (\sigma_\beta + \sigma_I + \sigma_R)\sqrt{2t \log\log t} - \sigma_R B_R(t)} \\ &- \frac{\gamma_k}{d_k^R - \frac{\sigma_R^2}{2} - \lambda} [e^{-(d_k^R - \frac{\sigma_R^2}{2} - \lambda)T(\sigma_\beta + \sigma_I + \sigma_R)\sqrt{2T \log\log T}} \\ &\times e^{-(d_k^R + \frac{\sigma_R^2}{2})t - \sigma_R B_R(t)}] \end{aligned} \tag{27}$$

therefore,

$$\limsup_{t \to \infty} \frac{1}{t} \log|w_k(t)| = -(d_k^R + \frac{\sigma_R^2}{2}) \vee -\lambda < 0 \tag{28}$$

Similarly, we can get the assertion for $u_k(t)$ as

$$\limsup_{t \to \infty} \frac{1}{t} \log|u_k(t)| = -\lambda \vee -(d_k^R + \frac{\sigma_R^2}{2}) \vee -d_k^S < 0 \tag{29}$$

In this way we proved that Eq. (15) is exponentially stable. According to the Oseledec multiplicative ergodic theorem [39], the necessary and sufficient condition for the almost sure asymptotic stability, of the trivial solution of the system is that the largest lyapunov exponent of the linearized system is negative. Therefore, we have the following results:

Theorem 3.1 Assume that $B = (\beta_{ij})_{n \times n}$ is irreducible.

(1) If $R_0^S < 1$, then the disease-free equilibrium P_0 is almost sure asymptotically stable, which means the disease will die out almost surely.

(2) If $R_0^S > 1$, then the disease-free equilibrium P_0 is unstable.

Remark 3.2 It is useful to observe that in either the classical deterministic model or the stochastic model, there is a threshold which reflect the

prevalent or extinction of the epidemic, but the thresholds are different between them, the stochastic threshold R_0^S is smaller than the deterministic one. In other words, the conditions for I(t) to become extinct in the SDE epidemic model are weaker than in the classical deterministic epidemic model. We give the following example illustrates this result more explicitly.

Example 3.3 For simplicity, let k = 2 and we choose the following system parameters

$A_1 = 100; A_2 = 300; d_1^S = 2; d_2^S = 3; d_1^I = 3; d_2^I = 5; d_1^R = 3;$

$d_2^R = 5; \beta_{11} = 0.1; \beta_{12} = 0.2; \beta_{21} = 0.3; \beta_{22} = 0.4; \varepsilon_1 = 1;$

$\varepsilon_2 = 1; \gamma_1 = 1; \gamma_2 = 1;$

so the stochastic multi-group SIR model in Eq. (6) becomes

$$\begin{cases} \mathrm{d}\, S_1 = (100 - 2S_1 - 0.1S_1I_1 - 0.2S_1I_2)\,\mathrm{d}\,t - \sigma_1 S_1 I_1\,\mathrm{d}\,B(t), \\ \mathrm{d}\, I_1 = (0.1S_1I_1 + 0.2S_1I_2 - 5E_1)\,\mathrm{d}\,t + \sigma_1 S_1 I_1\,\mathrm{d}\,B(t) - \sigma_I I_1\,\mathrm{d}\,B(t), \\ \mathrm{d}\, R_1 = (I_1 - 3R_1)\,\mathrm{d}\,t - \sigma_R R_1\,\mathrm{d}\,B(t), \\ \mathrm{d}\, S_2 = (300 - 3S_1 - 0.3S_2I_1 - 0.4S_2I_2)\,\mathrm{d}\,t - \sigma_2 S_2 I_2\,\mathrm{d}\,B(t), \\ \mathrm{d}\, I_2 = (0.3S_2I_1 + 0.4S_2I_2 - 7I_2)\,\mathrm{d}\,t + \sigma_2 S_2 I_2\,\mathrm{d}\,B(t) - \sigma_I I_2\,\mathrm{d}\,B(t), \\ \mathrm{d}\, R_1 = (I_2 - 5R_2)\,\mathrm{d}\,t - \sigma_R R_2\,\mathrm{d}\,B(t). \end{cases} \tag{30}$$

Clearly, if $\sigma_1 = \sigma_2 = \sigma_I = \sigma_R = 0$, Eq. (30) becomes to be the related deterministic multi-group SIR model and $R_0 > 1$ so $I_1(t)$, $I_2(t)$ will tend to their endemic equilibrium. In Fig. 1, we start our numerical simulation with $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $\sigma_I = \sigma_R = 0$, and the initial value are $I_1(0) = 10$, $I_2(0) = 20$. Noting that R_0^S<1, by Theorem 3.1, $I_1(t)$, $I_2(t)$ will tend to zero exponentially. Computer simulation in Fig.1 illustrates the extinction of disease.

Next we keep the parameter value and start our computer simulation at the initial value $I_1(0) = I_2(0) =$ 1,we gain the same results in Fig. 2.

If we decrease the environment intensity to $\sigma_1 =$ 0.02, σ_2= 0.01, $\sigma_I = \sigma_R = 0$ and starting from $I_1(0) = I_2(0) = 1$, which means R_0^S>1. From Theorem 3.1, the disease-free equilibrium will be unstable, result of one simulation run in Fig. (3) proves our results.

4. Stochastic Persistence

For a deterministic model, persistence is implied by showing the endemic equilibrium is a global attractor. But for our stochastic model Eq. (6), there is no endemic equilibrium.

In fact, the solution of our model is a process, if we can prove the process is positive recurrence relative to a positive domain, and then this conclusion implies the persistence of our stochastic model in stochastic version. Before proving the main result of this section, we give the concept of recurrence [40].

Definition 4.1 Let U be some bounded or unbounded domain, and we denote its complement U^c by U_1. A process $X(t)$ is said to be recurrence relative to the domain U (or U–recurrence) if it is regular and foe every $s, x \in U_1$.

$$p^{s,x}\{\tau_{U_1} < \infty\} = 1 \tag{31}$$

where, τ_{U_1} is the first exit time from U_1. A recurrent process with finite mean recurrence time is said to be positive recurrent.

The proof of positive recurrence result for the stochastic Eq. (6) is based upon the following lemma (see Theorem 3.12 in [41]).

Lemma 4.2 A necessary and sufficient condition for positive recurrence with respect to a domain $U = D \times l \subset R^r \times M$ is that for each $I \in$ M, there exists a nonnegative function $V(x, i)$:$D^c \rightarrow R$ such that $V(x, i)$ is twice continuously differentiable and that

$$\mathrm{LV}(x,i) = -1, (x,i) \in D^c \times M. \tag{32}$$

Following theorem is the positive recurrence result for the stochastic system in Eq. (6).

Theorem 4.3 Assume that $B = (\beta_{ij})_{n\times n}$ is irreducible and $R_0^S > 1$. If the random perturbation coefficient satisfies,

$$2(d_k^I + \varepsilon_k + \gamma_k) > \frac{(d_k^S + d_k^I + \varepsilon_k + \gamma_k)^2}{d_k^S} + \sigma_I^2, \qquad d_k^R > \sigma_R^2,$$

then our model Eq. (6) is positive recurrence relative to a positive domain, where $P^* = (S_1^*, I_1^*, R_1^*, \ldots, S_n^*, I_n^*, R_n^*)$ is the endemic equilibrium of the related deterministic system of Eq.(1).

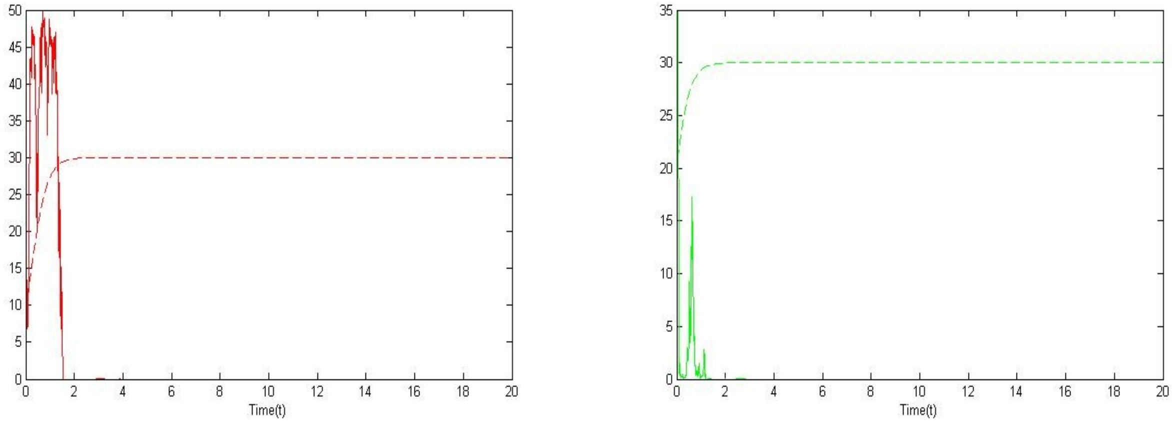


Fig. 1 Computer simulation of path $I_1(t)$, $I_2(t)$ for Eq. (30) and its corresponding deterministic model, using the EM method with step size 0.001, with initial value$I_1(0)$ = 10, $I_2(0)$ = 20. The full line express stochastic model's simulation, and the dotted line express the related deterministic model.

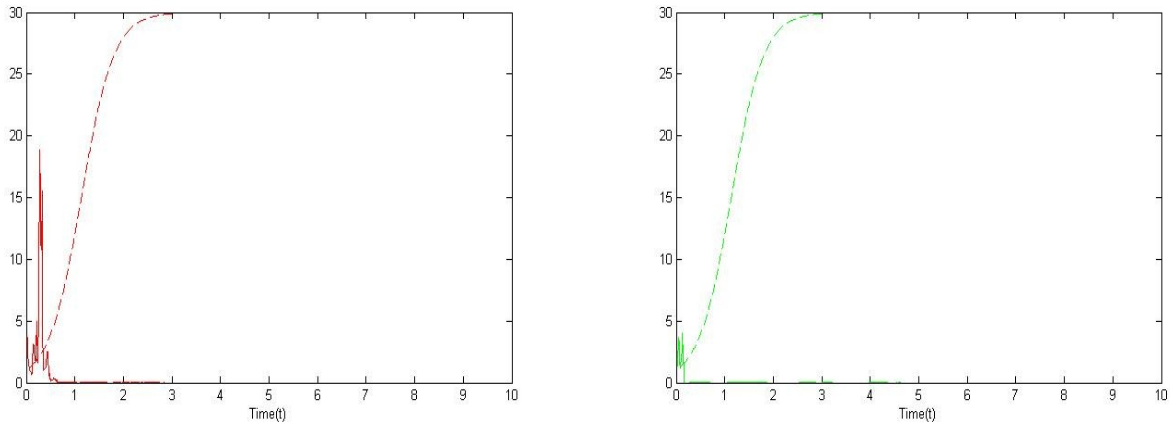


Fig. 2 Computer simulation of path $I_1(t)$, $I_2(t)$ for Eq. (30) and its corresponding deterministic model, using the EM method with step size 0.001, with initial value $I_1(0)$ = 1 = $I_2(0)$ = 1. The full line express stochastic model's simulation, and the dotted line express the related deterministic model.

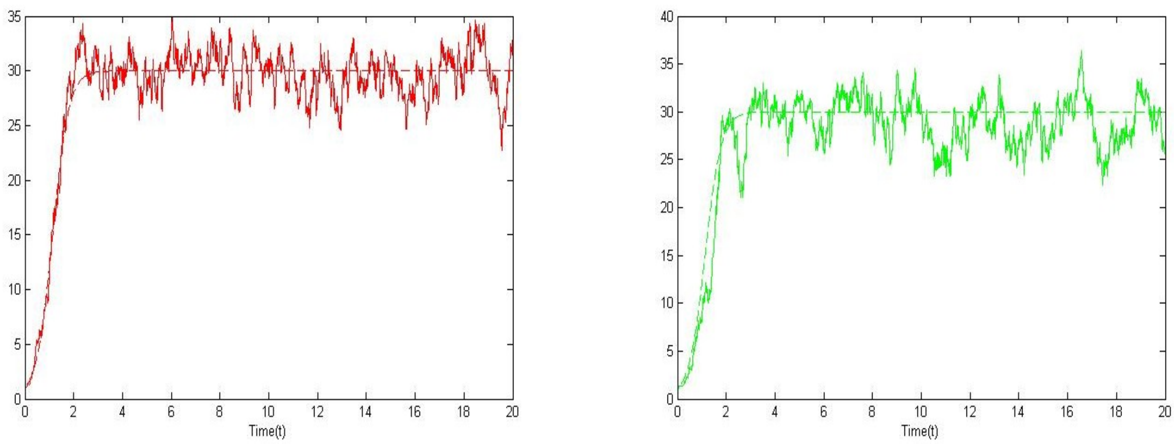


Fig. 3 Computer simulation of path $I_1(t)$, $I_2(t)$ for Eq. (30) and its corresponding deterministic model, using the EM method with step size 0.001, with initial value $I_1(0)$ =1, $I_2(0)$ = 1. The full line express stochastic model's simulation and the dotted line express the related deterministic model.

Proof. Since $1 < R_0^S < R_0$, there is an endemic equilibrium $P^* = (S_1^*, I_1^* R_1^*, \cdots, S_n^*, E_n^*, I_n^*)$ for the deterministic system of Eq. (15). We obtain the following equation

$$A_k = d_k^S S_k^* + \sum_{j=1}^{n} \beta_{kj} S_k^* I_j^*; \sum_{j=1}^{n} \beta_{kj} S_k^* I_j^* = (d_k^I + \varepsilon_k + \gamma_k) I_i; \gamma_k I_k = d_k^R R_k, \tag{33}$$

Set

$$\begin{aligned} V = \sum_{k=1}^{n} v_k (S_k - S_k^* \ln S_k + I_i - I_i^* \ln I_i) + a_k (S_k - S_k^* \\ + I_k - I_k^*)^2 + b_k (R_k - R_k^*)^2 \\ = V_1 + V_2 + V_3. \end{aligned} \tag{34}$$

Using the same method as that for the proof of Theorem 3.3 [7], we choose $\overline{\beta}_{kj} = \beta_{kj} S_k^* I_j^*, 1 \le k, j \le n$, $\overline{B} = (\overline{\beta}_{kj}), \{v_1, ..., v_n\}, v_k > 0$ such that $\overline{B} v = 0$. Applying $\hat{\text{Ito}}$ formula, we can calculate that

$$\begin{aligned} LV_1 &= \sum_{k=1}^{n} v_k [A_k - d_k^S S_k - \sum_{j=1}^{n} \beta_{kj} S_k I_j - A_k \frac{S_k^*}{S_k} + d_k^S S_k^* + \sum_{j=1}^{n} \beta_{kj} S_k^* I_j \\ &+ \frac{S_k^* \sigma_k^2 I_k^2}{2} + \sum_{j=1}^{n} \beta_{kj} S_k I_j - (d_k^I + \varepsilon_k + \gamma_k) I_k - \sum_{j=1}^{n} \beta_{kj} S_k I_j \frac{I_k^*}{I_k} + \\ &+ (d_k^I + \varepsilon_i + \gamma_k) I_k^* + \frac{I_k^* (\sigma_k^2 S_k^2 + \sigma_I^2)}{2}] \\ &= \sum_{k=1}^{n} v_k [d_k^S S_k^* (2 - \frac{S_k^*}{S_k} - \frac{S_k}{S_k^*}) + (\sum_{j=1}^{n} \beta_{kj} S_k^* I_j - (d_k^I + \gamma_k + \varepsilon_k) I_k \\ &+ (2 \sum_{j=1}^{n} \beta_{kj} S_k^* I_k^* - \sum_{j=1}^{n} \beta_{kj} I_j^* \frac{(S_k^*)^2}{S_k} - \sum_{j=1}^{n} \beta_{kj} S_k I_j \frac{I_k^*}{I_k}) + \\ &+ \frac{S_k^* \sigma_k^2}{2} I_k^2 + \frac{I_k^* \sigma_k^2}{2} S_k^2 + \frac{I_k^* \sigma_I^2}{2}] \\ &\le \sum_{k=1}^{n} v_k [(2 \sum_{j=1}^{n} \beta_{kj} S_k^* I_k^* - \sum_{j=1}^{n} \beta_{kj} I_j^* \frac{(S_k^*)^2}{S_k} - \sum_{j=1}^{n} \beta_{kj} S_k I_j \frac{I_k^*}{I_k}) \\ &+ \frac{S_k^2 \sigma_k^2}{2} I_k^2 + \frac{I_k^* \sigma_k^2}{2} S_k^2 + \frac{I_k^* \sigma_I^2}{2}] \\ &\le \sum_{k=1}^{n} v_k [\overline{\beta}_{kj} (2 - \frac{S_k^2}{S_k} - \frac{I_j S_k I_k^*}{I_k S_k^* I_j^*}) + \frac{S_k^* \sigma_k^2}{2} I_k^2 + \frac{I_k^* \sigma_k^2}{2} S_k^2 \\ &+ \frac{I_k^* \sigma_I^2}{2}] \end{aligned} \tag{35}$$

Calculating LV_2 we obtain

$$LV_2 = \sum_{k=1}^{n} a_k [2(S_k - S_k^* + I_k - I_k^*)(A_k - d_k^S S_k - (d_k^I - \varepsilon_k + \gamma_k) I_k) + \sigma_I^2 I_k^2] \tag{36}$$

Substituting Eq. (33) into (36) yields

$$\begin{aligned} LV_2 &= \sum_{k=1}^{n} a_k [2(S_k - S_k^* + I_k - I_k^*)(-d_k^S (S_k - S_k^*) - (d_k^I + \varepsilon_k + \gamma_k)(I_k - I_k^*)) + \sigma_I^2 I_k^2] \\ &= \sum_{k=1}^{n} a_k [-2 d_k^S (S_k - S_k^*)^2 - 2(d_k^I + \varepsilon_k + \gamma_k)(I_k - I_k^*)^2 + \sigma_I^2 I_k^2 - 2(d_k^S + d_k^I + \varepsilon_k + \gamma_k)(S_k - S_k^*)(I_k - I_k^*)] \\ &\le \sum_{k=1}^{n} a_k [-d_k^S (S_k - S_k^*)^2 - [2(d_k^I + \varepsilon_k + \gamma_k) - \frac{(d_k^S + d_k^I + \varepsilon_k + \gamma_k)^2}{d_k^S}](I_k - I_k^*)^2 + \sigma_I^2 I_k^2]. \end{aligned} \tag{37}$$

Similarly, we can calculate LV_3

$$\begin{aligned} LV_3 &= \sum_{k=1}^{n} b_k [2(R_k - R_k^*)(\gamma_k (I_k - I_k^*) - d_k^R (R_k - R_k^*)) + \sigma_R^2 R_k^2] \\ &= \sum_{k=1}^{n} b_k [2\gamma_k (I_k - I_k^*)(R_k - R_k^*) - 2 d_k^R (R_k - R_k^*)^2 + \sigma_R^2 R_k^2] \\ &\le \sum_{k=1}^{n} b_k [\frac{\gamma_k^2}{d_k^R} (I_k - I_k^*)^2 - d_k^R (R_k - R_k^*)^2 + \sigma_R^2 R_k^2]. \end{aligned} \tag{38}$$

Choose

$$m_k = 2(d_k^I + \varepsilon_k + \gamma_k) - \frac{(d_k^S + d_k^I + \varepsilon_k + \gamma_k)^2}{d_k^S}, b_k = \varepsilon, a_k = (\frac{v_k I_k^* \sigma_k^2}{2 d_k^S} + \varepsilon) \vee (\frac{\varepsilon \gamma_k^2}{(m_k - \sigma_I^2) d_k^R} + \varepsilon)$$

then

$$\begin{aligned} LV &\le \sum_{k=1}^{n} v_k \overline{\beta}_{kj} (2 - \frac{S_k^*}{S_k} - \frac{I_j S_k I_k^*}{I_k S_k^* I_j^*} + \frac{I_k^* \sigma_I^2}{2}) \\ &+ \sum_{k=1}^{n} [\frac{v_k I_k^* \sigma_k^2}{2} S_k^2 - (\frac{v_k I_k^* \sigma_k^2}{2} + \varepsilon)(S_k - S_k^*)^2] \\ &+ \sum_{k=1}^{n} [\frac{\varepsilon \gamma_k^2}{d_k^R} I_k^2 - (\frac{\varepsilon \gamma_k^2}{d_k^R} + \varepsilon)(I_k - I_k^*)^2] \\ &+ \sum_{k=1}^{n} [\varepsilon \sigma_R^2 R_k^2 - \varepsilon d_k^R (R_k - R_k^*)^2] \\ &= H(S_1, I_1, R_1, \ldots, S_n, I_n, R_n). \end{aligned} \tag{39}$$

Note that

$$2(d_k^I + \varepsilon_k + \gamma_k) > \frac{d_k^S + d_k^I + \varepsilon_k + \gamma_k)}{d_k^S} + \sigma_I^2, d_k^R > \sigma_R^2.$$

It is clearly that

$$\begin{aligned} \lim_{S_k, I_k, R_k \to 0} H(S_1, I_1, R_1, \ldots, S_n, I_n, R_n) = -\infty \\ \lim_{S_k, I_k, R_k \to \infty} H(S_1, I_1, R_1, \ldots, S_n, I_n, R_n) = -\infty \end{aligned} \tag{40}$$

So there exists a domain U lies entirely in R_+^{3n}. For $(S_1, I_1, R_1, ..., S_n, I_n, R_n) \in U_+^{3n}$, $LV < -M$, where M is a positive constant. According to Lemma 4.2, the

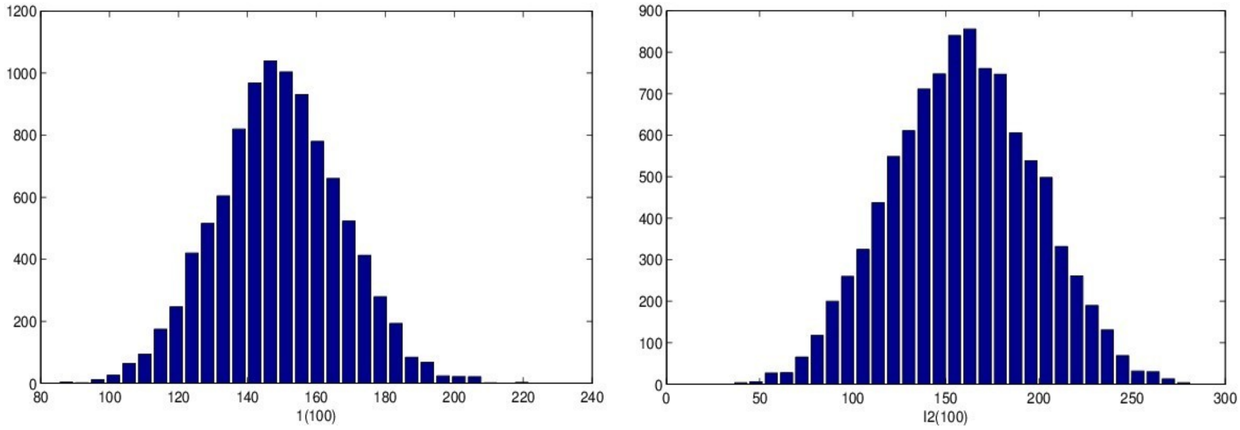


Fig. 4 Frequency histograms of path $I_1(t)$, $I_2(t)$ for Eq. (30) based on 10000 stochastic simulations for each population at time t = 100, using the EM method with step size 0.001,with initial value $I_1(0) = 1$, $I_2(0) = 1$.

solution of our model is positive recurrence related to the positive domain U. The proof is complete.

Example 5.4 To substantiate the analytic findings above, we provide numerical simulation results for the stochastic model Eq. (30). We also use the same parameters in Example 3.3, and let σ_1 = 0.02, σ_2 = 0.01,$\sigma_I = \sigma_R$ = 0. We have shown in Fig. 3 that $I_1(t)$, $I_2(t)$ will not tend to 0, Theorem 4.3 tell us that the solution will recurrence relative to a positive domain. Fig. 4 shows histograms of the approximate distribution of the infective classes.

5. Conclusions

In this paper, we have considered the general multi-group SIR epidemic model which in presence of multiplicative noise terms to understand the dynamics in presence of environmental driving forces. First we guarantee the existence and uniqueness of positive global solution of our stochastic epidemic model. Then we find out a sharp threshold which determines the extinction or persistence of disease. Specifically, If $R_0^S < 1$, then the disease-free equilibrium will be asymptotically stable which means the disease will die out, if $R_0^S > 1$, our model will positively recurrence to a positive domain which implies the persistence of our model. We need to point out that the sharp threshold we found is different from the related deterministic model's threshold, it smaller then the deterministic one. It makes sense in biological systems, because it means the environment perturbation increase the parameter values of the extinction of disease. After the theoretic proof, we give some numerical examples to illustrate our results.

Acknowledgments

This work was supported by the National Natural Science Foundation of China Grant 61273126, and the Natural Science Foundation of Guangdong Province Under Grants 10251064101000008 and S201210009675, the Fundamental Research Funds for the Central Universities 2012ZM0059, and Research Fund for the Doctoral Program of Higher Education of China under grant 20130172110027.

References

[1] W.O. Kermack, A.G. McKendrick, Contributions to the mathmatical theory of epidemics (Part 1), Proc. R. Sm. A 115 (1927) 700-721.

[2] B. Song, C. Castillo-Chavez, J.P. Aparicio, Tuberculosis models with fast and slow dynamics: The role of close and casual contacts, Math Biosci 180 (1-2) (2002) 187-205.

[3] M.Y. Li, J.R. Graef, L. Wang, J. Karsai, Global dynamics of a SEIR model with varying total population size, Math Biosci 160 (2) (1999) 191-213.

[4] M.Y. Li, H.L. Smith, L.Wang, Global dynamics of an SEIR epidemic model with vertical transmission, SIAM, J. Appl. Math. 62 (1) (2001) 58-69.

[5] L. Wen, X. Yang, Global stability of a delayed SIRS model with temporary immunity, Chaos Solitons Fract 38 (1) (2008) 221-226.

[6] Hsu Sze-Bi, Lih-Ing W. Roeger, The final size of a SARS

epidemic model without quaran-tine, J. Math. Anal. Appl. 333 (1) (2007) 557-566.

[7] H.B. Guo, M.Y. Li, Z.S. Shuai, Global stability of the endemic equilibrium of multigroup SIR epidemic models, Canada Apl. Math. Q.14 (1) (2006) 259-284.

[8] J. Liu, Y. Zhou, Global stability of an SIRS epidemic model with transport-related infection, Chaos Solitons Fract 40 (1) (2009) 145-158.

[9] H-F. Huo, Z-P. Ma, Dynamics of a delayed epidemic model with non-monotonic incidence rate, Commun Nonlinear Sci Numer Simul 15 (2) (2010) 459-468.

[10] C.C. McCluskey, Complete global stability for an SIR epidemic model with delay—Distributed or discrete, Nonlinear Anal RWA 11 (1) (2010) 55-59.

[11] A. Lajmanovish, J.A. York, A deterministic model for gonorrhea in a nonhomogeneous population, Math. Biosci 28 (1976) 221-236.

[12] H.B. Guo, M.Y. Li, Global dynamics of a staged progression model for infectious diseases, Math. Biosci 3 (2006) 513-525.

[13] E.Beretta, V.Capasso, Global stability results for a multi-group SIR epidemic model, World Scientific, Singapore,1986, pp. 317-342.

[14] Hsu Sze-Bi, Lih-Ing W. Roeger, The final size of a SARS epidemic model without quaran-tine, J. Math. Anal. Appl. 333(2007) 557-566.

[15] H.R Thieme, Methmatics in population biology, Princeton University Press, Princeton, 2003.

[16] I.G Lauko, Stability of disease free sets in epidemic models, Math. Comput. Modeling 43(2006) 1357-1366.

[17] D. Xiao, S. Ruan, Global stability of an epidemic model with nonmonotone SIR epidemic models, Math. Biosci. 208 (2007) 419-429.

[18] S. Ruoyan, Global stability of the endemic equilibrium of multigroup SIR models with nonlinear incidence, Comput. and Math. with Appl. 60(2010) 2286-2291.

[19] M.Y. Li, Z. Shuai, Global-stability problem for coupled systems of differential equations on networks, J. Diff. Equa. 248 (2010) 1-20.

[20] R.Z. Khasminskii, F.C. Klerbaner, Long term behavior of solution of the Lotka-Volterra system under small random perturbations, Ann. Appl. Probab. 11 (2001) 952-963.

[21] A. Bahar, X. Mao, Stochastic delay Lotka-Volterra model, J. Math. Anal. Appl. 292 (2004) 364-380.

[22] D.Q. Jiang, N.Z. Shi, A note on nonautonomous Logistic equation with random perturbation, J. Math. Anal. Appl. 303 (2005) 164-172.

[23] D.Q. Jiang, N.Z. Shi, X.Y. Li, Global stability and stochastic permanence of a non-autonomous logistic equation with random perturbation, J. Math. Anal. Appl. 340 (1) (2008) 588-597.

[24] L. Imhof, S. Walcher, Exclusion and Persistence in deterministic and stochastic chemostat models, J. Differ Equ. 217 (1) (2005) 26-53.

[25] M. Carletti, On the stability properties of a stochastic model for phage-bacteria interaction in open matine environment, Math. Biosci. 175 (2002) 17-131.

[26] N. Dalal, D. Greenhalgh, X.R. Mao, A stochastic model for internal HIV dynamics, J. Math. Anal. Appl. 341 (2008) 1084-1101.

[27] N. Dalal, D. Greenhalgh, X.R. Mao, A stochastic model of AIDS and condom use, J. Math. Anal. Appl. 325 (2007) 36-53.

[28] X. Mao, G. Marion, E. Renshaw, Asymptotic behaviour of the stochastic Lotka-Volterra model, J. Math. Anal. Appl. 287 (2003) 141-156.

[29] X. Mao, G. Marion, E. Renshaw, Environmental Brownian noise suppresses explosion in population dynamics, Stoch. Process Appl. 97 (2002) 95-110.

[30] E. Tornatore, S.M. Buccellato, P. Vetro, Stability of a stochastic SIR system, Phys A: Statistical Mechanics and its Applications 354 (2005) 111-126.

[31] J.J. Yu, D.Q. Jiang, N.Z. Shi, Global stability of two-group SIR model with random perturbation, J. Math. Anal. Appl. 360 (2009) 235-244.

[32] C.Y. Ji, D.Q. Jiang, Q.S. Yang, N.S. Shi, Dynamics of a multigroup SIR epidemic model with stochastic perturbation, Automatica 48 (2012) 121-131.

[33] H. Liu, Q.S. Yang, D.Q. Jiang, The asymptotic behavior of stochastically perturbed DI SIR epidemic models with saturated incidences, Automatica 48 (2012) 820-825.

[34] C.Y. Ji, D.Q. Jiang, N.Z. Shi, The behavior of an SIR epidemic model with stochastic pertur-bation, Stochastic Anal. Appl. 30 (2012) 755-773.

[35] A. Gray, D. Greenhalgh, L. Hu, X. Mao, J. Pan, A stochastic differential equation SIS epidemic model, Siam. J. Appl. Math. 71 (3) (2011) 876-902.

[36] Horst R. Thieme, Uniform persistence and permanence for non-autonomous semiows in population biology, Mathematical Biosciencess 166 (2000) 173-201.

[37] M. Liu, K. Wang, Q. Wu, Survival analysis of stochastic competitive models in a polluted environment and stochastic competitive exclusion principle, Bull Math. Biol. 73 (2011) 1969-2012.

[38] X.R. Mao, Stochastic Differential Equation and Application, Horwood Pub., Glasgow, 1997.

[39] V.I. Oseledec, A multiplication ergodic theorem: Lyapunov charactoristic numbers for dynamical system, Trans. Moscow Math. Soc. 19 (1968) 197-231.

[40] R,Z. Hasminskii, Stochastic stability of differential equations, Alphen aan den Rijn, Sijthoff & Noordhoff, The Netherlands, 1980.

[41] C. Zhu, G. Yin, Asymptotic properties of hybrid diffusion systems, SIAM, J. Control Optim, 49 (4) (2007) 1155-1179.

Journal of Control Science and Engineering 1 (2013) 23-37

"Intelligent" Knowledge Reuse for Complex Logistics Projects: An Application of Ontology-Driven and Case-Based Reasoning

Stephan Zelewski, Martin Kowalski and Daniel Bergenrodt

Institute for Production and Industrial Information Management, University of Duisburg-Essen, Essen 45141, Germany

Abstract: The application potential of ontology-driven and CBR (case-based reasoning) is demonstrated for the business knowledge management particularly with respect to the reuse of knowledge of experience concerning logistics projects. The relevance of poorly structured, qualitative and especially in natural language represented knowledge is outlined for purposes of knowledge management, particularly with respect to the management of project-related knowledge. It is elucidated how this kind of knowledge can be preprocessed and reused with the support of a computer. At first, the technique of CBR is outlined in its basic fundamentals. Thereupon, it will be shown how the technique of ontologies can be used for the computer-supported processing of knowledge represented in natural language and integrated in computer-assisted CBR systems. A simple application example illustrates how ontology-driven and CBR can be used in practice within the reuse of project-related knowledge. Finally, it will be addressed which further need for research exists in principle.

Key words: Case-based reasoning, knowledge management, knowledge of experience, knowledge reuse, logistics projects, ontologies, project management.

1 Problems of Knowledge Management in the Context of Complex Logistics Projects

This paper deals with the computer-assisted management of complex logistics projects as they exist, for example, within Supply Chain Management or within large international transport projects. The management of such projects stretches mostly to the management of project-related knowledge [1-3], here referred to as project-knowledge for short. Therefore, it is sometimes directly referred to as "project knowledge management" [4].

This knowledge-based approach to project management suggests the idea that a project is a typical knowledge-intensive task [5]. In addition, "empirical evidence" can be adduced claiming that in practice for the management of complex projects it usually turns out to be the most promising approach for management not to plan a new project from scratch each time. Instead, practical knowledge that is already available in a company from older, already realised—preferably similar—projects should be "intelligently" reused [3, 6].

However, this knowledge-based approach to project management in business reality is subject to some serious difficulties. They are further outlined as the real problems of business knowledge management in relation to the management of project-related knowledge.

The business practice points out that knowledge on projects can only be represented to a small part in a well-structured, usually quantitative manner or at least in formal language. This sort of knowledge is mostly referred to as explicit knowledge. From now on, this term will be used continuously on behalf of the integrability of established linguistic habits, because it

Corresponding author: Martin Kowalski, Dipl.-Inf., research fields: computer science, logistics. E-mail: martin.kowalski@pim.uni-due.de.

is in wide use in the field of knowledge management to distinguish between explicit (or easily codifiable) and implicit (or tacit) knowledge [7, 8]. However, from a scientific point of view this is a problematic manner of speaking, as implicit knowledge can be explicated, e.g., like Nonaka et al. [8, 9] demonstrated with their concept of knowledge conversion and—in a "more dynamic" regard even more developed—with their spiral of knowledge [8, 9].

Explicit project knowledge extends to a large part to quantitative aspects. This includes, for example, the execution times of single processes or work packages, the cost of process or work package designs as well as on the capacity requirements for this process or work package designs, particularly with regard to the necessary personnel, equipment and transportation capacity. Equally relevant are sequence relationships between individual exporting operations or work packages, which cannot be quantitatively but precisely expressed using precedence relationships in formal language.

This explicit project knowledge can be organised and archived in a well-structured manner, and also reused, especially with the help of project databases. For a first or repeated application of this explicit project knowledge, sophisticated business planning techniques are available, like the network planning technique. They make it possible to manage the explicit project knowledge in terms of "hard" economic goals of temporal or monetary nature, such as the observance of delivery dates or the minimisation of the project costs, with no fundamental difficulties. For supporting knowledge management, with regard to this explicit project knowledge and with respect to the "hard" project goals, there is well-engineered project management software.

The quantitative explicit project knowledge is not subject of the present paper, though. The reason is that it is frequently not nearly sufficient for a successful project management. Usually, it is a large number of qualitative influences mostly expressible only in a way of natural language that affects the success of a project. These influences, that are often referred to as critical success factors from a business perspective, play an important role, especially in the context of the strategic project management. Instead, they cannot be listed here comprehensively for brevity's sake, the elaborations in Ref.. [10].

Two examples have to suffice for clarification. On the one hand, the qualifications of the assigned project staff often have a very strong influence on the success of the project. These qualifications concerning the personnel are also known as competencies in terms of empowering knowledge [11]. On the other hand, the success of a project can considerably be influenced by the reliability, especially the sanctity of contracts of actors participating in a project—individuals, groups of individuals (e.g., assembly crews) or entire companies. Moreover, the success of a project from a business perspective cannot only be assessed on the basis of the id. "hard" goals, time and cost goals in particular, but also in terms of "soft" goals. The "soft" goals include additional qualitative aspects. This concerns, for example, the quality of project implementation from the customer perspective (project purchaser), who is able to influence significantly future acquisitions of further project orders as "reputation drivers". The compliance of social expectations towards such vague concepts as CSR (corporate social responsibility) [12] or good governance [13] need to be also considered.

Quantifying or at the very least formalizing genuine qualitative aspects like competencies, reliability, project quality or even CSR respectively good governance would be regarded as plain nonsense by any experienced project managers. Most of the knowledge experienced project managers possess on such influences or goals of qualitative kind is not saved in project databases, because they are not able to do so. Instead, they usually keep it in "their minds" as implicit or—more distinctive in this context from the perspective of the authors—as tacit knowledge [8, 14-16].

The reasons for not entering this implicit respectively tacit knowledge of experience into the project databases in explicit form can be multifarious. At least three reasons turn out in business practice to be significant barriers for explicating. Firstly, it could simply appear too time-consuming to the project managers to "codify" their knowledge of experience in a form that can be governed by project databases. Secondly, it is possible that project managers regard their knowledge—often acquired through many years of professional activity—as some sort of knowledge for the sake of action or control which ensures their indispensability and therefore, should not come to the attention of third parties. Thirdly, it is imaginable that project managers know how to use their knowledge of experience intuitively, but they are unable to express this knowledge in a form of natural language; this is the case of tacit knowledge in the narrow sense of the word.

As long as qualitative project knowledge stays locked up in form of implicit or tacit knowledge in the "heads" of project managers, it can be applied by the projects managers as knowledge owners to be sure and also reused personally. But it turns out to be impossible to have this qualitative project knowledge mutually—be it for the first time or yet again—used by basically all members of a company that plan, process and control a project. For this mutual, especially repeated, use of knowledge it is necessary to explicate and communicate this qualitative project knowledge.

The explication of qualitative project knowledge by means of quantification or at least formalization is usually ruled out for the already stated reasons. Instead, qualitative project knowledge, which possesses the character of the management's knowledge of experience from former projects, is mainly represented in a manner of natural language. The project knowledge represented in a qualitative manner and in natural language can barely be found in conventional databases (although this is generally not impossible), but mostly in the form of documents in natural language [2, 6, 17]. Typical examples for this are the "debriefings", "lessons learnt" and "post-project reviews" [14].

Such documents are made by the significant project staff—especially project managers—after the processing of a project in order to save the knowledge on important success as well as failure factors for reuse with future projects. The documents include some well-structured information in the document header at best. For instance, information on customers, project sector, project duration and project costs are part of this. The largest part of the document, however, consists of largely unstructured "continuous text". That is why the knowledge of experience of already realised projects, which is codified in such documents, is referred to as semi-structured or ill-structured knowledge. The authors prefer the latter term because the amount of unstructured or barely structured knowledge predominated in documents like this by far.

The major challenge of project knowledge management entails the processing of ill-structured and qualitative knowledge of experience mostly represented in natural language of already realised old projects and making it available to the project staff in a way that it can be successfully reused in new projects (desideratum of the reuse of project knowledge). This complies with the general recommendation made within business knowledge management time and again to reuse already existing knowledge [7, 18, 19].

The recommendation for a systematic reuse of project-related knowledge of experience is often not implemented despite apparent economic advantages in business practice. The intended reuse of knowledge usually succeeds only with well-structured and usually quantitative factual knowledge ("know what"), which can be, e.g., easily saved with the help of conventional database systems and recalled straightforwardly for the purpose of reuse. Though, with ill-structured and mostly qualitative knowledge predominantly

represented in natural language, which embraces skills in terms of action-based knowledge ("know-how") and "everyday theories" for the pragmatic explanation of processes and concepts ("know-why"), the recommended reuse of knowledge frequently encounters considerable barriers [20, 21].

The barriers concerning the intended reuse of knowledge are mainly based on three reasons. Firstly, it is difficult to save and recall qualitative, predominantly natural linguistically represented knowledge using conventional database systems (the problem of qualitative knowledge). Instead, in the best case it is filed as explicit knowledge in natural language documents, i.e., free-formatted "texts". In the worst case, it is even only "locked up" as implicit knowledge "in the minds" of some professionals (experts). Secondly, in the event of implementing a new project, it is very intricate to assess whether it is similar enough to older, already realised projects so that it can be, at least partially, worked out with the knowledge of experience gained from the implementation of older projects (the problem of sufficient problem similarity). Thirdly, the knowledge of experience, that is available in a company on the matter of successfully, or deficiently, treating previous projects, is so extensive that in business practice it is barely possible to realise a systematic reuse of that knowledge without computer support. The usage of computers in the field of knowledge management often fails in reality because of the need of being able to process qualitative and predominantly natural linguistically represented knowledge (the problem of computer support).

2 Innovative Contributions by the Research of Artificial Intelligence as "Enabler" for Knowledge Reuse within Project Management

2.1 Overview

Up to now, conventional papers on business knowledge management have not contributed a convincing solution concept that demonstrated how the previously sketched problems of qualitative knowledge, sufficient problem similarity and computer support can be solved or at least minimized from the point of view of business practice within the range of the reuse of project knowledge. The problems of sufficient problem similarity and computer support are not, at least not with the preferable substantial depth, elaborated in business papers concerning knowledge management. The problem of qualitative knowledge is admittedly frequently addressed from the perspective of "implicit" and "tacit" knowledge, but it adheres mostly to the less operational recommendation to manage this knowledge in "discursive" or "interactive" processes. Indications as to how the processes of the reuse of qualitative knowledge can be effectively supported by the intensive usage of computers cannot be found in the relevant specialized literature for knowledge management, at least from the business perspective.

On the part of artificial intelligence research, an interesting "combined" approach to a problem is provided. On the one hand, it is the technique[1] of CBR (case-based reasoning). It shows a "natural", direct connection to the latter problems of sufficient problem similarity and computer support. On the other hand, the CBR technique can be "enriched" with the help of ontologies in a way that it shows a promising problem-solving potential for the first-mentioned problem of qualitative knowledge. In this paper, it is outlined how the combination of ontology-driven and CBR can be applied in order to realize processes of the reuse of knowledge within the management of complex logistics projects in an innovative and computer-supported way.

The technique of CBR serves the reuse of knowledge by transferring the knowledge of experience gained during the treatment of earlier

[1]A technique is here understood as a complex of generic model structures, of methods for the model solution as well as of instruments for the computer-assisted application of models and methods.

realized projects on the treatment of new, "sufficiently" similar projects. The technique of ontologies puts itself forward to formally structure and process, e.g., on the level of "semantics", qualitative knowledge especially represented in natural language. Both techniques agree on using automatic information processing systems ("computers") in order to fulfill their special purpose because of their rooting in the research of artificial intelligence.

Up to now, both techniques of CBR and of ontologies have been researched independently of each other to a large extent due to their different fields of application. From a business point of view they have barely been used for, even in the course of knowledge management, solving practical business problems so far. That is why the innovative approach of the here presented paper entails combining CBR and ontologies with each other so that they can be used together—"synergetically"—for the support of the reuse of knowledge in project management.

The underlying research takes part in the course of the joint research project OrGoLo (organizational innovations via good governance in logistics networks) which is part of the German efficiency cluster "Efficiency Cluster Logistics Ruhr", supported by the BMBF (German Ministry for Education and Research).

2.2 Problem Solving Based on Analogies with the Help of Case-Based Reasoning

One of the most promising techniques of reusing knowledge of experience from old, already realised projects for new projects is the CBR. The CBR technique has been developed within artificial intelligence research in particular [22-27].

CBR can be regarded as a special technique of "intelligent" knowledge processing which is based on the principle of analogical problem solving. At this kind of knowledge processing, a complex problem is not decomposed in a conventional, analytical way into simple-to-handle sub-problems in order to eventually synthesise a solution for the original encompassing problem from the solutions of the sub-problems. Instead, in order to solve a new — usually equally complex problem, it is searched for experiences with old, already solved and, if possible, similar problems. This knowledge of experience on already known solutions is transferred to the new, yet to be solved problem and, if necessary, adapted to its specificities.

The epistemic justification of transferring solution-based knowledge on old problems to new problems is, on the one hand, caused by the similarities of old and new problems as well as, on the other hand, the epistemic core belief that similar solutions are convenient for similar problems (analogy principle).

The actual validity of the analogy principle or at least its universal applicability can be critically raised to question. For example, insights from the theory of "deterministic chaos" give reasons to doubt that similar causes (problems) always lead to similar effects (solutions). Instead, already slight, e.g., accidental variations on constituting the causes of a phenomenon can lead to very different manifestations of this phenomenon—like in the so called "butterfly effect". Complications of this kind are disregarded on principle in the case of problem solving based on analogies.

In the context of CBR, problems are generally thematised as cases. This is also true for the here discussed projects, because the implementation of a project can be understood as a special manifestation of the general problem concept: the "generic" problem of project management entails from a business point of view the realization of a material aim (the object of the project specification) in a way that formal goals predefined by the decision making units can be fulfilled at the best (extremity goals)—or at least satisfactorily (satisfying goals). This specialised terminology is used here for the sake of compatibility with established literature.

Each case can be specified within the CBR in a particularly comprehensive manner of representation

via three characteristic components: the case description (the problem description), the result (the problem solution) and the evaluation (the evaluation of the problem solution) [28]. Projects can be seen as a special type of such cases. Therefore, in the following the terms "problems", "cases" and "projects" can be regarded as synonyms.

The knowledge of experience about already realised, old projects is stored in a project knowledge base— or synonymous: A case base. Such project knowledge bases are often also referred to as "knowledge repositories" [3, 17, 19]. A project knowledge base contains the descriptions, results and evaluations for all former realised projects.

The typical process of reusing the project knowledge for the management of new projects can be divided into the "4-R-phases" of the so-called CBR cycle based on Ref. [25] in particular. These four cycle phases: "retrieve", "reuse", "revise" and "retain" are illustrated in Fig. 1.

The description of a new project presents the problem which initiates the CBR cycle. The project description is used to search for at least one sufficiently similar and—if there exist several sufficiently similar projects—at least one most similar old project in the knowledge base ("retrieve"). If no sufficiently similar old project can be found in the project knowledge base, the CBR cycle is terminated and another problem solving technique has to be used. The case-based reasoning is therefore no "certain" problem solving technique that guarantees to find an acceptable, i.e. sufficiently similar, solution candidate for every new problem—thus for every new project.

Otherwise, if at least one sufficiently similar old project exists and if at least one most similar old project has been chosen from the set of sufficiently similar old projects, each most similar old project chosen has to be adjusted to the regarded, new project respectively ("reuse"), because the new project normally does not correspond exactly to the old projects chosen.

The quality of the adjustment result can be evaluated in view of additional user requirements of a project management system, like with regard to the integrity and practicability of the result, and—in case of insufficient evaluation results—it can be modified under certain circumstances ("revise"). Failure to fulfill these requirements results in the unsuccessful termination of the CBR cycle once again.

If all user requirements are fulfilled, the description, the result and the evaluation of the new project are conflated in order to save them as new project-related knowledge of experience, i.e., as a learned new case, in the project knowledge base for future reuse of knowledge ("retain").

In order to implement this pattern of the CBR cycle for the knowledge reuse of business project managements in a user friendly way, the development and application tool jCOLIBRI has been chosen by the joint research project OrGoLo as open source software [29, 30]. Above all, there were three reasons for choosing this software tool. Firstly, it is equipped with a broad selection of measurements of similarity, which can be applied not only on quantitative (numerically represented) but also on qualitative objects (represented in natural language). Secondly, it can be combined with a common software tool for editing ontologies (Protégé) without any difficulties. Thirdly, the software tool jCOLIBRI can be flexibly adapted to special wants of the users with the help of self-provided Java routines. The latter mentioned option has been used in order to integrate extended ontology-based measures of similarity. This will be discussed later.

In the center of the development work were real projects of practice partners which concerned the management of international supply chains. This included suitable packaging for the shipping of components of a polar research station and their carriage by sea from Duisburg to the Antarctic [31] and also the multi-modal transport of DIY-products from their factory within China to their distribution

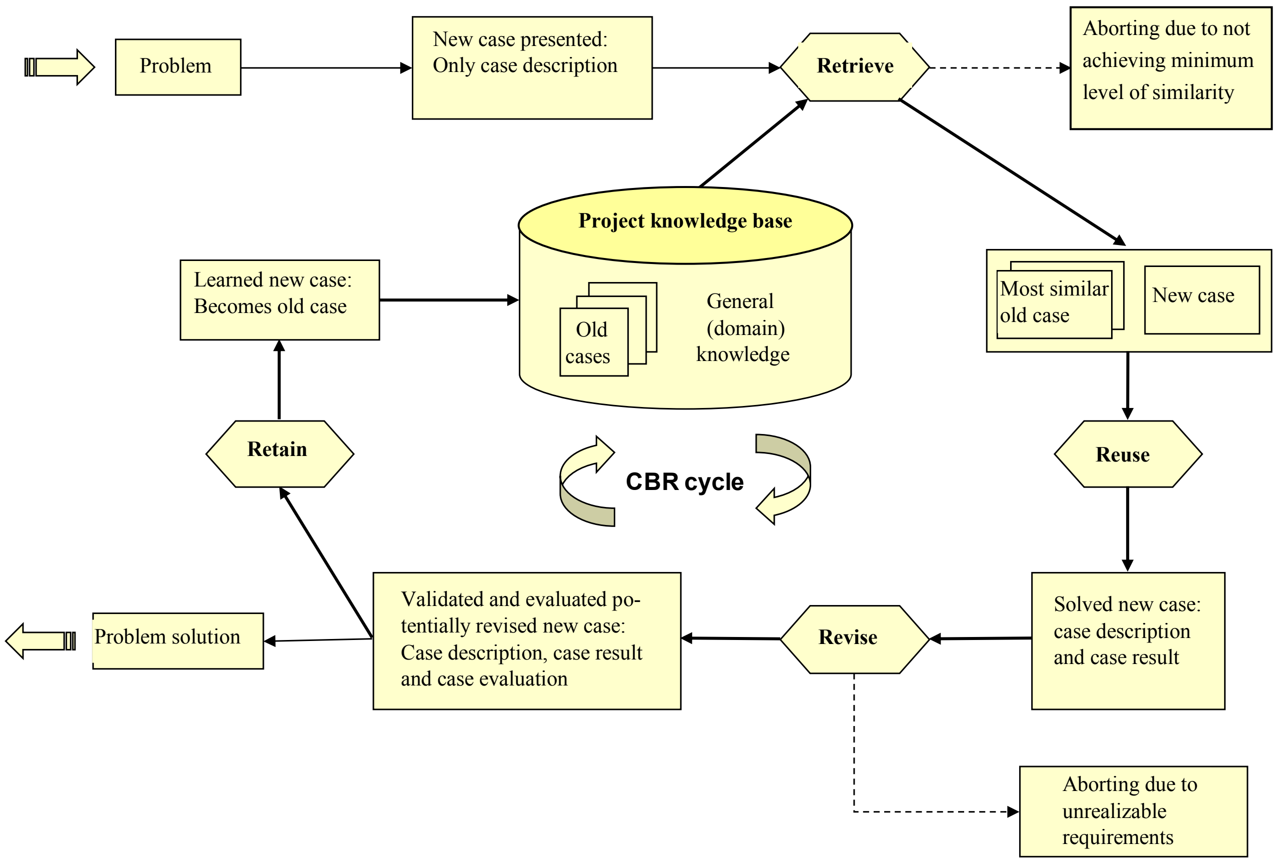


Fig. 1 The CBR cycle (Aamodt and Plaza [25], augmented in Kowalski et al. [10]).

centers of a well-known trade chain in Germany [32].

A user of the CBR tool describes his new logistics project by means of project characteristics that have been identified as characteristics "crucial for success" by the cooperation of experienced practice partners of the joint research project OrGoLo. These project characteristics entail in the here regarded example information on the transport relation (departure and target place, possibly also meeting with intermediate goals), the transport carrier to be used and the chosen means of transportation, the kind of freight to be transported (HS-position and dangerous goods class), as well as the packaging, amount, weight and size of the freight to be transported, the aimed at delivery date, the (monetary) size of an order as well as legal particularities (letter of credit and international commercial terms). The project characteristics are partially of quantitative nature. For example, amount, weight and size of the freight to be transported are counted among them. A plethora of qualitative, natural linguistically expressed project characteristics have to be also considered, like the place of the transport relation, the kind of the freight and packaging as well as terms from the legal sphere.

The user can assign different significances to the individual project characteristics with respect to the currently regarded, new logistics project in order to determine which project characteristics influence the search after (at least) one most similar and at once sufficiently similar old project in the project knowledge base particularly strongly or which project characteristics have only little effect on this search.

Further insights into the functionality of the CBR tool "SCM Project Recommender" convey the papers of Zelewski et al. [33] and Kowalski et al. [34].

In a second stage of development of the joint research project OrGoLo, it is planned to extend the functionality of the prototype "SCM Project

Recommender" particularly with regard to the reuse phase in order to adjust via computer, the found most similar old projects to the particularities of the regarded new logistics project respectively. For these case adjustments knowledge of experienced project managers is necessary. This knowledge of experience usually is "locked up in their minds" and it is only hard to explicate. The acquisition of this adjustment knowledge and its representation in a computer-processable form—like in the form of adjustment rules—is a big challenge for the development of a practical CBR tool. First preliminary work on explicit codifying—such adjustment rules for case-based reasoning—is already available from other research work [34, 35]. But it still needs to be revised for the domain of logistics projects.

2.3 Representing Project Knowledge by Ontologies

In the field of the project management, a wealth of project-related knowledge exists. It is knowledge of experience about already realised, old projects in particular that could be reused "in principle" for the planning, implementation and management of new projects, but is actually barely used. This reusable knowledge mainly exists in the form of a large number of documents of realised old projects with ill-structured, qualitative knowledge predominantly represented in natural language. The documents have usually been created with text processing software within the involved companies. Therefore, they are generally available on computer-based databases or document servers. As these on computer-based systems saved documents contain considerable knowledge of experience on already realised projects, the systems are here referred to as project knowledge bases.

Although promising preconditions exist because of the existence of such project knowledge bases to support knowledge intensive business processes of the project management—particularly processes of the reuse of project-related knowledge—with instruments of the automatic information processing. However, currently available computer-assisted project management systems are restricted to the ability to find documents with knowledge of experience about old projects in a knowledge base on the basis of given descriptions of new projects, if the new project description and the old project documents prove to be similar on a purely syntactic level of text sequences ("strings"). Therefore, the search in a project knowledge base for document about similar, already realised projects happens with the help of a search item (or its combinations) as a simple "string matching". A content-addressed and hence semantic search for reusable knowledge does not happen in such a way. Hence, it is not surprising that the few attempts to use CBR for the purpose of project management that have been noted so far [6, 22, 36-38] turn out not to be successful—at least with regard to the requirements of business practice on such computer-assisted project management systems.

Because of the aforementioned reasons, a big challenge for computer-assisted knowledge management lies in the processing of knowledge of experience about old, already realised projects, that especially exists in the form of ill-structured documents represented in natural language with a large amount of qualitative project knowledge, in a way that it can be easily accessed via computer and reused with regard to new projects. This challenge can be paraphrased—a bit constricted as to content, but bold and simple on that account—as "dilemma of natural language": On the one hand, the natural language of knowledge representation is often necessary to present qualitative project knowledge in a manner of expression that is rich in substance and familiar to the business users of computer systems; On the other hand, the usage of computer systems requires a knowledge representation in formal language. Furthermore, the measurement of similarities between project descriptions requires quantitative similarity measures with regard to the project characteristics to be compared.

This is aggravated by the fact that the documents which contain the biggest part of knowledge of experience about old, already realised projects are not only ill-structured and written in natural language, but also turn out to be heterogeneous concerning their presupposed terminologies. The terminological heterogeneity can be hardly avoided in complex, especially international logistics projects, because a wide range of actors (individuals, companies, governmental and non-governmental organizations, as well as even software systems) is involved that are used to express their thoughts in different company, organizational or software specific “dialects” as well as in different national languages.

Ontologies [37, 39-41] provide a promising approach to overcome the previously outlined difficulties as well as to process and reuse natural linguistically represented, especially qualitative and terminologically heterogeneous project knowledge with the support of a computer. With the help of ontologies, it is particularly possible to “measure” the semantic distances and thus the similarities between terms in natural language with which the qualitative knowledge about different projects is represented in the documents of a project knowledge base.

Due to the appropriate brevity at this point, it cannot be given a lecture on the scientific discussion about an “adequate” understanding of ontologies. Instead, an understanding of ontologies is taken as a basis that follows the frequently cited ontology definition by Gruber [41], but tries to avoid some deficiencies of this approach to a definition [6]: An ontology is an explicit specification in formal language of those linguistic means of expression that are considered as necessary for the construction of representational models of a commonly used conceptualization of real phenomena according to one of the several actors. In the course of this, the conceptualization extends to those real phenomena which are regarded by the actors as observable or imaginable in a subject and goal dependent restricted real world situation and which are used or needed for the communication between the actors.

The development of ontologies for business relevant domains is presently still indicative of considerable deficits. The ontologies that have been presented in this context until now usually stem from authors that can be included with the computer sciences and engineering. Their ontologies indicate without difficulty that little understanding exists for the business terminology as well as for the subtle linguistic nuances and distinctions that are common in business discussions. Especially for the domain of complex, mostly international logistics projects do currently not exist which prove to be terminologically acceptable from the business perspective. Therefore, a fundamental goal of development of the joint research project OrGoLo is to construct an ontology which caters for the above mentioned application field and fulfills the standards of business terminologies.

It is a special skill or even “virtuosity” to compare the project specific characteristics of qualitative, i.e., non-numerical attributes of projects on a quantitative similarity scale with the help of ontology. First approaches at solving this demanding problem are already available [22, 42-44, 46]. Beyond that, remarkable research approaches exist in which the here proposed combination of CBR on the one hand and ontologies on the other hand has already been considered [22, 29, 45-47].

An essential aim of the joint research projects OrGoLo is to develop special ontologies for the application field of complex, especially international logistics projects and to integrate them into the prototype “SCM Project Recommender” for the computer-supported CBR. Especially a suggestion from Beißel [22] to measure similarities between projects mostly described in natural language with the help of “path lengths” in graphically represented ontologies is implemented in the prototype “SCM Project Recommender” and therefore accessed for practical use.

3 An Application Example

In the following, it is outlined by means of a realistic example from Beißel [22]—how an ontology for mostly qualitative knowledge about projects can be used for the implementation of information and communication techniques in order to identify a similar already realised project in a CBR system and to reuse the knowledge of experience about this for the management of a new project. Within the joint project OrGoLo, the underlying conceptual attempt for ontology-driven and case-based reasoning is soon to be transferred on the structurally very similar management of knowledge about complex logistics projects.

The example concerns the new project to implement a patch management with WSUS (windows server update services) for a client in order to close security gaps on his computer system with the operating system Windows XP through installing small software packages ("patches") as well as remedy other small deficiencies. For this new project WSUS, the anticipated costs of the project implementation are to be estimated in order to be able to submit a proposal for the client. The project knowledge base contains the comprehensive information on over ten already realised, former implementation projects. They are real cases from business practice. By means of the software prototype for ontology-driven and CBR, an old project among others has been identified in the project knowledge base that turns out to be very similar to the new project. In this old project "Online-Banking", a web application has to be implemented on a server in a client company that is operated under the operating system SUSE Linux.

The whole similarity calculation for both regarded projects cannot be depicted here for the sake of brevity. Instead, it will be only outlined in an exemplary way how a partial project similarity can be determined with regard to a relational project attribute with qualitative expressions of characteristics by means of a domain specific ontology. It is the project characteristic "application software" which specifies that the application software WSUS is to be installed in the new project, whereas in the old project a "web application" had to be installed as an application software.

For both instances "WSUS" and "web application" of the category "application software", it is first determined by means of the underlying ontology, from which a segment is depicted in the following figure, that these instances belong to two different sub-categories "standard application" or alternatively "individual application". But the two sub-categories at least agree on both attributes "application name" and "application field". Thus, both instances can be at least compared in their common attributes in a first step.

By means of further reflections, that cannot be explained here in detail for brevity's sake, both instances are assigned only a very slight similarity of 0.1 with regard to their manifestations "WSUS" and "web application" of the attribute "application name". However, their similarity with respect to their manifestations "internet" and "network" of the attribute "application field" is significantly greater with a value of 0.6. Since both attributes "application name" and "application field" are rated as equally important by a user with respect to their relevance for cost assessment, a value of $(0.5 \times 0.1) + (0.5 \times 0.6) = 0.35$ arises for the weighted similarity of both instances relating to the project attribute "application software" and regarding their agreeing attributes (but not manifestations of attributes).

However, this similarity value is only temporary, because the knowledge that both instances "WSUS" and "web application" do not belong to a joint category has not been taken into account, yet, duringthe similarity determination. Only instances of the same category can be compared directly because of their similarity. For entities that belong to different categories, an ontology-based correction of

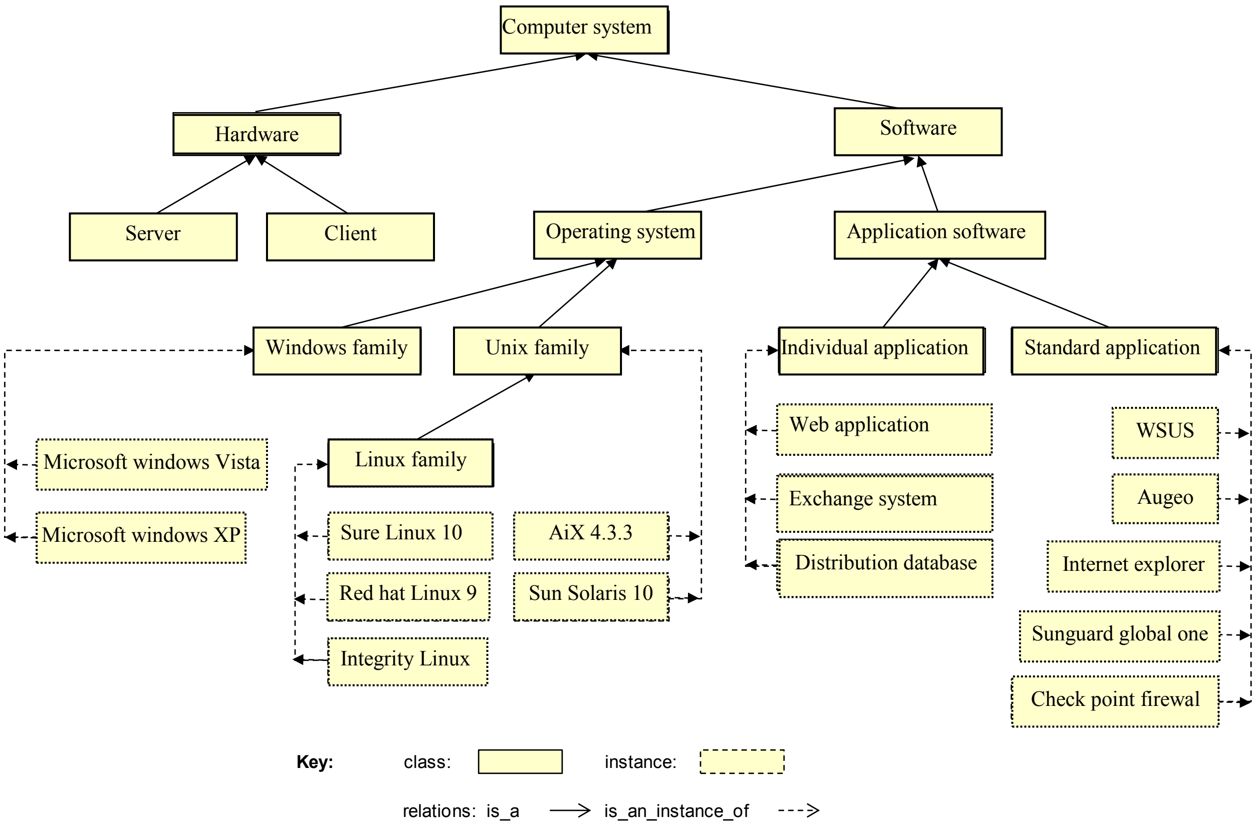


Fig. 2 Part of an ontology for information and communication technology implementation projects (referring to [22]).

the similarity values has to be conducted. This correction captures the scope of similarity between categories that are located in different "semantic positions" in the underlying ontology. On the one hand, this ontology-based correction of similarity values presents an especially difficult task. On the other hand, this correction justifies the special attraction and "surplus" of the combination of case-based reasoning and ontologies.

The similarity between two categories which are located in different "semantic positions" in the underlying ontology is determined by the relation of two attribute numbers. The first attribute number is the amount of all attributes which the two categories share. In the example here, these are the two attributes "application name" and "application field" for the categories "standard application" and "individual application". Since they belong to the two categories "standard application" and "individual application", the number of agreeing attributes of both categories is $2 \times 2 = 4$ in total. The second attribute number is the amount of all attributes which both regarded categories have available in total. These are altogether 7 attributes in the here regarded example. Thus, the relation of both aforementioned attribute numbers amounts to 4/7.

In addition to the ontology-driven correction, it is taken into account how remote two instances, which belongs to different categories in the underlying ontology, are maximally from that super category to which both different categories initially belong as sub-categories in the category hierarchy of the ontology. This is according to Fig. 2, the category "application software". Both instances "WSUS" and "web application" are 2 edges remote from the ontology so that the maximal distance of the instances of the initially superordinate super-category also amounts to 2. However, it is to be considered that the similarity between instances from different categories is even smaller the greater the distance of the

instances in the underlying ontology turns out to be. Therefore, in addition to the similarity measurement in an ontology, the reciprocal value for the maximal number of edges between the instances and the initially superordinate super-category has to be also used. This amounts to the additional correction value of 1/2.

Altogether, the following arises as a result: the partial similarity between the new and the old project with respect to the relational project attribute "application software" is composed of two components. On the one hand, it is the weighted similarity of both instances "WSUS" and "web application" regarding their agreeing attributes with the value 0.35. On the other hand, the ontology-driven correction of this similarity measurement for the similarity between the two categories is added to which both instances belong and which are located in different "semantic positions" in the underlying ontology. The correction value is here $(4/7) \times (1/2) \approx 0.29$. A value of $0.35 \times 0.29 \approx 0.10$ results from this for the partial similarity between the new and the old implementation project regarding the relational project attribute "application software".

The partial project similarities of all other project attributes are to be determined in the same way. In doing so, the complicated ontology-driven correction of a similarity value can be foregone if the two instances that specify the regarded new and old project with respect to a project attribute influencing the cost belong to the same category in the underlying ontology, and therefore, located in the same "semantic position". The sum of the partial similarity values for all project attributes finally amounts to the similarity value for the two compared implementation projects for information and communication techniques.

In the previously outlined way, it is possible to ascertain the similarity of projects of any kind that are characterized through numerous non-numerical, i.e., qualitative and usually natural linguistically formulated project attributes. This specifically applies to knowledge about complex logistics projects which are in the focus of the joint research project OrGoLo. In this way, a considerably greater amount of knowledge of experience about already realised projects can be used for the management of new logistics projects than it has been possible so far on the basis of purely quantitative similarity measurements in CBR without recourse to ontologies. Hence, computer-assisted estimations of the costs for a demanding logistics project in preparation for the participation in the competition for a new project can be realized on a far more extensive project knowledge basis and thus tendentially also more reliably than it has been until now with purely quantitative, especially statistical estimation techniques.

4. Summary and Outlook

The here presented research and transfer work for the reuse of knowledge of experience about complex logistics projects is based on a promising and innovative approach that is, however, also in need of enhancement. This approach combines ontology-driven and CBR for the purpose of business knowledge management. On the one hand, it is hereby for the first time possible to reuse extensively available qualitative, especially natural linguistically represented knowledge with the help of a computer for the management of new logistics projects. On the other hand, the practicability of this approach in the daily routine of business depends on whether some of the desiderata open from the perspective of research are fulfilled. The additional need for research especially ranges over five problematic issues.

Firstly, there is an urgent need for ontologies that are tailored for business application fields, like the here regarded management of complex, especially international logistics projects, and that exactly consider business terminology. Secondly, translation aids—also based on ontologies—are necessary that arbitrate between the heterogeneous terminologies of different actors that e.g. cooperate in such logistics

projects. Thirdly, considerable know—how deficiencies persist in the field of CBR regarding a vast number of configuration—worthy independent variables. This concerns, for example, the selection of attributes for the description of projects, the determination of attribute specific similarity scales as well as the number of most similar old projects (with sufficient project similarity), that should be used for the construction of a proposal for processing a new project. Fourthly, it proves to be a big problem to specify explicit rules or methods for the task of adapting the reused knowledge about old, already realised projects to the idiosyncrasies of very similar, but not identical, new projects. Fifthly, criteria, with which help the quality of proposals for processing new projects can be evaluated, are largely unexplored.

The previously mentioned research problems that are currently still awaiting a satisfying solution are to be worked on in greater detail within the joint research project OrGoLo.

Acknowledgements

This contribution presents results from the joint research project OrGoLo. This project is supported by the BMBF under the sign "01IC10L20A". The authors are grateful for the support.

References

[1] M. Görög, Translating single project management knowledge to project programs, Project Management Journal 42 (2) (2011) 17–31.

[2] D.J. Sage, A.R.J. Dainty, N.J. Brookes, Who reads the project file? Exploring the power effects of knowledge tools in construction project management, Construction Management and Economics 28 (6) (2010) 629–639.

[3] P. Arora, D. Owens, D. Khazanchi, A pattern-based tool for knowledge management in virtual projects, The IUP Journal of Knowledge Management 8 (3) (2010) 60–80.

[4] S. Gasik, A model of project knowledge management, Project Management Journal 42 (3) (2011) 23–44.

[5] K.U. Koskinen, Knowledge management to improve project communication and implementation, Project Management Journal 35 (1) (2004) 13–19.

[6] S. Zelewski, A. Bruns, M. Kowalski, Ontologies for Guaranteeing the Interoperability in E-Business: A Business Economics Point of View, in: E. Kajan, F.J. Dorloff, I. Bedini (Eds.), Handbook of Research on E-Business Standards and Protocols: Documents, Data and Advanced Web Technologies, IGI Global, Hershey, 2012, pp. 154–184.

[7] F. Lehner, Knowledge Management: Basics, Methods and Technical Support, 4th ed., Carl Hanser, Munich, 2012. (in German)

[8] I. Nonaka, E. von Krogh, Tacit knowledge and knowledge conversion: controversy and advancement in organizational knowledge creation theory, Organization Science 20 (3) (2009) 635–652.

[9] I. Nonaka, H. Takeuchi, The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation, Oxford University Press, Oxford, 1995.

[10] M. Kowalski, H. Klüpfel, S. Zelewski, D. Bergenrodt, A. Saur, Integration of Case-Based and Ontology-Based Reasoning for the Intelligent Reuse of Project-Related Nnowledge, in: U. Clausen, M. ten Hompel, M. Klumpp (Eds.), Efficiency and Logistics, Lecture Notes in Logistics, Vol. 1, Springer, Heidelberg, Berlin, 2013, pp. 289–299.

[11] H.T. Ingason, H.I. Jónasson, Contemporary Knowledge and Skill Requirements in Project Management, Project Management Journal 40 (2) (2009) 59–69.

[12] C. Geßner, L. Heidbrink, V. Kammel, A. Kölle, M. Kreuels, N. Meyer, et al, Integrated Corporate Social Responsibility Management in Logistics Networks (CoReLo), in: U. Clausen, M. ten Hompel, M. Klumpp (Eds.), Efficiency and Logistics, Lecture Notes in Logistics, Vol. 1, Springer, Heidelberg, Berlin, 2013, pp. 7–18.

[13] H. Lautenschläger, M. Lautenschläger, Good Governance in Global Supply Chains from Eight Perspectives, in: U. Clausen, M. ten Hompel, M. Klumpp (Eds.), Efficiency and Logistics, Lecture Notes in Logistics, Vol. 1, Springer, Heidelberg, Berlin, 2013, pp. 19–29.

[14] K. Goffin, U. Koners, Tacit knowledge, lessons learnt, and new product development, Journal of Product Innovation Management 28 (2) (2011) 300–318.

[15] J. Perraton, I. Tarrant, What does tacit knowledge actually explain?, Journal of Economic Methodology 14 (3) (2007) 353–370.

[16] M. Polanyi, The Tacit Dimension, Doubleday, New York, 1966.

[17] P. Ribino, A. Augello, G.L. Re, S. Gaglio, A knowledge management and decision support model for enterprises, Advances in Decision Sciences 15 (2011) 1–16.

[18] L.F. Lin, W.Y. Zhang, Y.C. Öou, C.Y. Chu, M. Cai, Developing manufacturing ontologies for knowledge reuse in distributed manufacturing environment,

International Journal of Production Research 49 (2) (2011) 343–359.

[19] M.L. Markus, Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success, Journal of Management Information Systems 18 (1) (2001) 57–93.

[20] P. Kenning, M. Blut, Barrieres of Knowledge Management from a Theoretical View, in: S. Zelewski, D. Ahlert, P. Kenning, R. Schütte (Eds.), Knowledge Management in Sales and Service Networks, Gabler, Wiesbaden, 2005, pp. 19–30. (in German)

[21] R. Schütte, Barrieres of knowledge management from a practical view, in: S. Zelewski, D. Ahlert, P. Kenning, R. Schütte (Eds.), Knowledge Management in Sales and Service Networks, Gabler, Wiesbaden, 2005, pp. 31–43. (in German)

[22] S. Beißel, Ontology-based case-based reasoning: Development and assessment of semantic similarity indicators for the reuse of project knowledge represented in natural language, doctoral dissertation, University of Duisburg-Essen, Gabler, Wiesbaden, 2011. (in German)

[23] I. Bichindaritz, Prototypical cases for retrieval, reuse, and knowledge maintenance in biomedical case-based reasoning, Computational Intelligence 25 (3) (2009) 214–234.

[24] Y. Avramenko, A. Kraslawski, Case based design: Applications in process engineering. Springer, Berlin, 2008.

[25] A. Aamodt, E. Plaza, Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, AICOM (AI Communications) 7 (1) (1994) 39–59.

[26] J. Kolodner Case-Based Reasoning, Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[27] C. Riesbeck, R. Schank, Inside Case-Based Reasoning, Lawrence Erlbaum Associates Publishers, Hillsdale, NJ, 1989.

[28] K. Kurbel, P. Dornhoff, Cost estimation for software development projects with the help of case-based knowledge processing, Zeitschrift für Betriebswirtschaft 63 (10) (1993) 1047–1065. (in German)

[29] J. De Miguel, L. Plaza, B. Díaz-Agudo, ColibriCook: A CBR System for Ontology-Based Recipe Retrieval and Adaption, in: M. Schaaf (Ed.) ECCBR 2008: The 9th European Conference on Case-Based Reasoning, Workshop Proceedings, Tharax, Hildesheim, Sep. 1-4, 2008, pp. 199–208.

[30] B. Díaz-Agudo, P. González-Calero, J.A. Recio-García, A. Sánchez-Ruiz-Granados, Building CBR systems with jCOLIBRI, Science of Computer Programming (69) (2007) 68–75.

[31] M. Kowalski, M. Heffe, Construction of a Case for Case-based Reasoning Using the Example of the Project "Polarstation" of the Duisport Packing Logistics, OrGoLo Project report No. 14, Institute for Production and Industrial Information Management, University of Duisburg-Essen, In-house publisher, Essen, 2012. (in German)

[32] M. Kowalski, C. Thomas, Description of an International Logistics Project for an Ontology-based Case-based Reasoning System, OrGoLo Project Report No. 17, Institute for Production and Industrial Information Management, University of Duisburg-Essen, In-house publisher, Essen, 2013. (in German)

[33] S. Zelewski, M. Kowalski, D. Bergenrodt, H. Klüpfel, Reuse of Knowledge in Logistics Projects Using Case-based Reasoning, in: F. Lehner, N. Amende, N. Fteimi (Eds.), Conference Proceedings of the 7th Conference on Professional Knowledge Management, Gito, Berlin, 2013, pp. 74–95. (in German)

[34] M. Kowalski, S. Zelewski, N. Günes, T. Kühn, Cost estimations for the reactivation of passive railways sidings: A new method for cost estimations with the help of CBR (case-based reasoning) based on the reuse of historical project knowledge, EI-Der Eisenbahningenieur 62 (6) (2011) 49–54. (in German)

[35] S. Zelewski, M. Kowalski, T. Kühn, Cost Estimation for the Reactivation of Railways Sidings with the Help of Case-Based Reasoning, in: S. Zelewski, S. Jene (Eds.): Cooperation between Railway Undertakings: Basics-Concepts-Practical Applications, Vol. 1. Logos, Berlin, 2011, pp. 489–503. (in German)

[36] J. Chou, Web-Based CBR system applied to early cost budgeting for pavement maintenance project, Expert Systems with Applications 36 (2) (2009) 2947–2960.

[37] Y. Lin, M. Xie, T. Goh, A study of project selection and feature weighting for analogy based software cost estimation, Journal of Systems and Software 82 (2) (2009) 241–252.

[38] S. Dogan, D. Arditi, M. Asce, M. Günaydin, Determining attribute weights in a CBR model for early cost prediction of structural systems, Journal of Constructions Engineering and Management 132 (10) (2006) 1092–1098.

[39] W.Y. Zhang, J.W. Yin, L.F. Lin, T.H. Zhu, Towards a general ontology of multidisciplinary collaborative design for Semantic Web applications, International Journal of Computer Integrated Manufacturing 22 (12) (2009) 1144–1153.

[40] N. Guarino, Understanding, building and using ontologies, International Journal of Human-Computer Studies 46 (2/3) (1997) 293–310.

[41] T.R. Gruber, Toward principles for the design of ontologies used for knowledge sharing, International

Journal of Human-Computer Studies 43 (1995) 907–928.

[42] A.A. Assali, D. Lenne, B. Debray, Heterogeneity in Ontological CBR System, in: S. Montani, L.C. Jain (Eds.) Successful Case-Based Reasoning Application, SCI 305. Springer, Heidelberg, Berlin, 2010, pp. 97–116.

[43] T. Roth-Berghofer, B. Adrian, A. Dengel, Case Acquisition from Text: Ontology-Based Information Extraction with SCOOBIE for My CBR, in: I. Bichindaritz, S. Montani (Eds.), Case-Based Reasoning Research and Development, 18th International Conference, ICCBR 2010, Lecture Notes in Artificial Intelligence, Vol. 6176, Springer, Heidelberg, Berlin, 2010, pp. 451–464.

[44] A. Maedche, S. Staab, Measuring Similarity Between Ontologies, in: A. Gomez-Pérez, V.R. Benjamins (Eds.), Knowledge Engineering and Knowledge Management. Lecture Notes in Computer Science, Vol. 2473, Springer, Heidelberg, Berlin, 2002, pp. 251–263.

[45] S. Manzano, S. Ontanon, E. Plaza, A Case-Based Approach to Mutual Adaptation of Taxonomic Ontologies, in: B. Diaz-Agudo, I. Watson (Eds.), Case-Based Reasoning Research and Development, 20th International Conference, ICCBR 2012. Lecture Notes in Artificial Intelligence, Vol. 7466, Springer, Heidelberg, Berlin, 2012, pp. 226–224.

[46] V. Dufour-Lussier, J. Lieber, E. Nauer, Y. Toussaint, Improving Case Retrieval by Enrichment of the Domain Ontology, in: A. Ram, N. Wiratunga (Eds.) Case-Based Reasoning Research and Development, 19th International Conference, ICCBR 2011, Lecture Notes in Artificial Intelligence, Vol. 6880, Springer, Heidelberg, Berlin, 2011, pp. 62–76.

[47] A.A. Assali, D. Lenne, B. Debray, Case Retrieval in Ontology-Based CBR Systems, in: B. Mertsching, M. Hund, Z. Aziz (Eds.), KI 2009, Advances in Artificial Intelligence, Vol. 5803, Springer, Heidelberg, Berlin, 2009, pp. 564–571.

Journal of Control Science and Engineering 1(2013) 38-45

Evaluation of Traffic Improvement Options Using Traffic Simulation

Bashar H. Al-Omari[1], Fouad A. Gharaybeh[1] and Ahmad H. Alomari[2]

1. Jordan University of Science & Technology, Irbid 22110, Jordan

2. Civil Engineering Department, University of Central Florida, Florida 32816, USA

Abstract: This research aimed at evaluating traffic improvement options for urban street networks using traffic simulation. Traffic, geometric and control data were collected from a selected study area which consisted of eight major at-grade intersections in Irbid city, Jordan, and the traffic simulation software (CORSIM) was used to evaluate the proposed traffic improvement options. The results of the network level analysis showed that the scenario of two rings of one-way streets provided the highest average travel speed (40% increase) and the lowest average traffic delay (47% reduction) as compared to the existing condition. It was found that a network with moderate intersections traffic volumes performs better with roundabouts than with signals. Replacing the network roundabouts by coordinated and optimized traffic signals provided more improvement in the average travel speed than the average traffic delay.

Key words: Delay, speed, network, congestion, simulation.

1. Introduction

Traffic agencies in Jordan are providing solutions for mitigating traffic congestion, however, most of their improvements are applied at individual intersections resulting, in many cases, in shifting the problems to other locations. It would be more efficient to consider comprehensive solutions on the network level taking in consideration the interactions and effects of different intersections. However, this is not possible to be evaluated in the real world due to cost and safety concerns. Several comprehensive improvement alternatives may be evaluated by the traffic simulation software as the most practical and effective tools for quantifying the limitations and benefits of different improvement options.

Irbid city was taken as an example for main cities in Jordan to evaluate selected practical traffic management scenarios using computer simulation. Irbid city has the second largest metropolitan population in Jordan after the capital city, Amman with a population of around 650,000, and is located about 70 km to the north of Amman. This research aimed at using CORSIM as a tool for evaluating the proposed traffic improvement options in reducing traffic congestion in urban areas.

2. Literature Review

Different methods are usually followed for mitigating traffic congestion in urban areas. Schrank and Lomax [1] found that traffic signal coordination reduced the arterials delay by about 1.5%. Meyer [2] listed the major congestion reduction strategies along arterials and local streets considering the roadway design (one-way streets, reversible traffic lanes, arterial access management etc.) and operations (traffic signal improvements, turn prohibitions, improved traffic control devices etc.). Arnold [3] concluded that congestion reducing measures by

Corresponding author: Bashar Hani Salem Al-Omari, associate professor, research fields: traffic engineering operation, safety, simulation, control, and behaviour. Email: bahomari@just.edu.jo.

managing the existing supply are usually rated above average in effectiveness and below average in cost and ease of implementation.

Several studies have focused on the advantages of one-way over two-way street systems. Dixit [4] found that networks with increased fraction of one-way streets have reduced driver's perception of likelihood to crash. Enustun [5] found that converting some state trunk lines through cities from two-way to one-way has increased the average travel speed up to 28.2 mph in Lansing. Hocherman and Bar-Ziv [6] reported that converting two-way streets to one-way streets has resulted in a decrease in traffic accidents by 20%-30%. Wright et al. [7] showed that different combinations of one-way streets and turn restrictions can be used to minimize the frequency and severity of traffic conflicts.

On the other hand, Wang et al. [8] found that two-way street systems produce lower emissions than one-way street systems during off-peak periods, but not necessarily during peak periods. Gayah and Daganzo [9] found that the two-way operation with banned left turns has higher trip-serving capacity than the one-way operation. Hawkins et al. [10] evaluated the transportation and safety impacts of converting some existing one-way roadways into two-way operations within downtown CBD of Des Moines due to the belief from retail businesses that two-way operations will attract more customers. Meredith and Prem [11] evaluated the feasibility of eliminating one-way restrictions in the CBD of Kansas city and identified the negative impacts (increase in vehicle hours-of-travel and potential for more accidents) and the positive impacts (decrease in vehicle miles-of-travel).

Dbindsa and Spiller [12] have explained that the recurring localized bottlenecks occur when the approaching traffic rate exceeds the departing traffic rate due to decision points and/or physical constraints. They recommended using the highway capacity manual procedures for analyzing bottlenecks at localized areas, and the simulation procedures for expanded study areas where priority is given to improving the entire system considering different improvement strategies. Yun and Park [13] have shown that the microscopic simulation software like CORSIM outperform the macroscopic analysis software like SYNCHRO in evaluating the coordinated actuated traffic signal networks.

Lin et al. [14] have used traffic simulation and recommended that the traffic congestion at a selected major intersection can be reduced by establishing an alternate route to bypass that intersection reducing the average delay to more than 30%. Chaudhry et al. [15] have utilized traffic simulation among other techniques to develop guidelines for mitigating congestion in traffic signal systems.

CORSIM was used by several researchers to compare alternative control and design strategies for traffic intersections, corridors and networks. It was used by Lin et al. [16] for evaluating the proposed model-based controllers with different criteria to decrease network traffic congestion, by Hale [17] to test the HCM2010 trajectory analysis procedures, by Hummer [18] to study the double-wide design as an inexpensive way to increase capacity of signalized intersections on four-lane highways, by Knapp et al. [19]to study the feasibility of converting the number of lanes on an urban minor arterial from four to three, by Yu and Meyer [20] to develop a downtown traffic circulation plan for the city of Topeka- Kansas, and by Lu et al. [21] to evaluate the safety and operational impacts of different alternative left-turn treatments from driveways and side streets.

3. Methodology

The methodology of this research can be summarized as follows:

(1) Data collection: The collected data included geometric data (the number of lanes, lane widths etc.), traffic data (traffic volumes at links and intersections, traffic mix etc.), and control data (signal timings, phasing etc.).

(2) Simulation model development: The study area traffic network was modeled on CORSIM considering the existing geometric, traffic and control conditions. Part of the collected field data was used for the model calibration and validation. CORSIM is a microscopic and stochastic simulation software which applies time step simulation to model individual vehicle movement theories on a second-by-second basis based on car following theory, gap acceptance, lane changing, and other models for describing the interactions among the driver, vehicle, road and environment [22]. Fig. 1 shows the modeling of the study area on CORSIM with zoomed animation at one intersection.

(3) Analysis of existing traffic condition: CORSIM was used to evaluate the existing network on both individual level (for each major intersection) and network level (considering the whole study area network) using the evaluation criteria based on delay and speed.

(4) Developing traffic improvement alternatives: Based on the results of the analysis and utilizing the input from Irbid municipality, a group of practical alternatives were developed to reduce the traffic congestion in the study area.

(5) Evaluating alternatives and choosing the best alternative: The traffic simulation was used to evaluate different alternatives and the best alternative was selected and recommended.

4. Data Collection and Reduction

The selected study area consists of eight major intersections around Yarmouk University in Irbid city, Jordan as shown in Fig. 1. All geometric, traffic and control data were collected by Alomari [23] with the help of Irbid municipality during the spring of 2010. Tables 1 and 2 show the main intersection geometric elements.

Automatic traffic counts were conducted on the major network links to identify the peak periods. It was found that all intersections had the same a.m. peak period; however, they had different noon and p.m. peak periods. So, it was decided to adopt the a.m. peak period for the rest of the analysis.

Manual traffic counts were conducted to get the volumes of turning movements at all network major intersections, considering the main turning movements

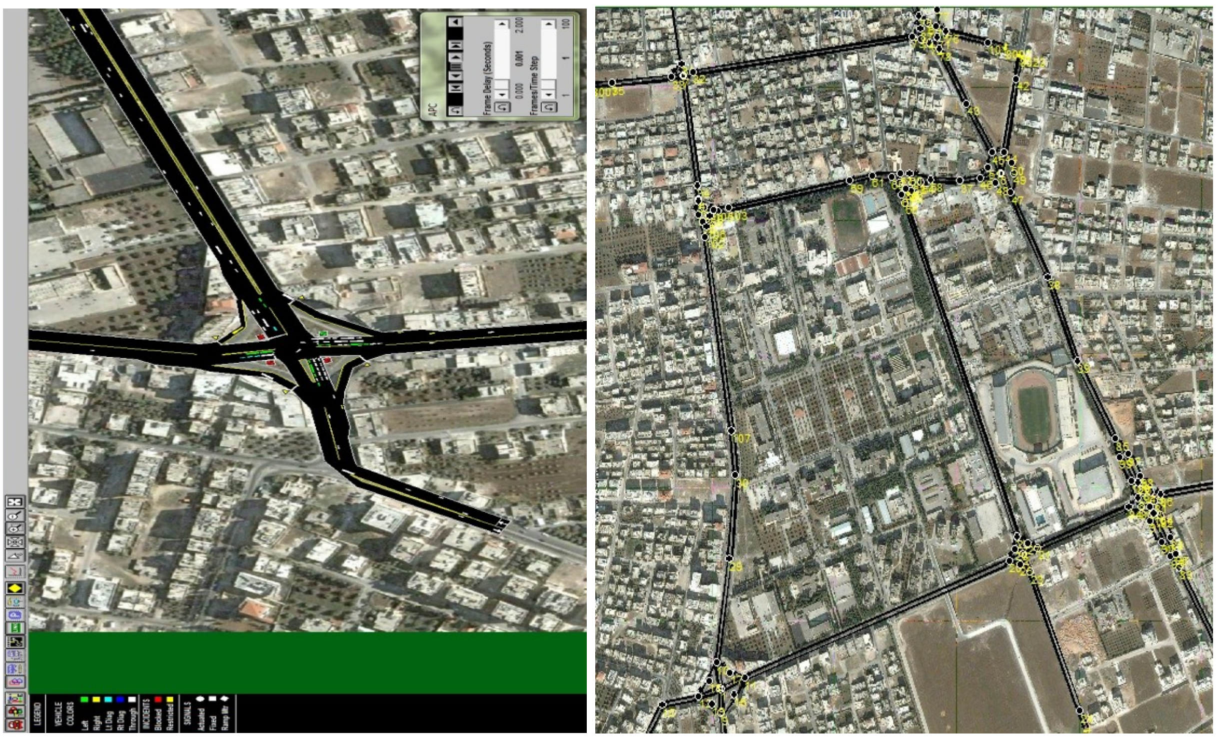


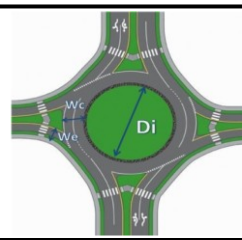
Fig. 1 Modeling the study network on CORSIM (with zoomed animation at one intersection).

Table 1 Intersections characteristics.

Intersection	Control type	Number of lanes				Lane width (m)				Number of approaches
		EB	WB	NB	SB	EB	WB	NB	SB	
T1	Traffic signal	3	4	3	3	3.4	3.3	3.3	3.7	4
T2	Traffic signal	4	4	2	2	3.3	3.2	3.7	3.4	4
T3	Roundabout	2	2	2	2	3.8	3.8	3.8	3.8	4
T4	Roundabout	-	3	3	3	-	3.1	3.1	3.1	3
T5	Roundabout	3	3	3	-	3.8	3.1	3.5	-	3
T6	Traffic signal	4	2	4	4	3.3	3.7	3.3	3.3	4
T7	Traffic signal	2	2	2	3	4	3.6	3.6	3.3	4
T8	Roundabout	2	2	3	2	3.8	3.7	3.5	3.6	4

Table 2 Roundabouts characteristics.

Intersection	[1]Di (m)	[2]We (m)				[3]Wc (m)
		EB	WB	NB	SB	
T3	48	7.6	7.6	7.6	7.6	11.5
T4	16	-	9	9	9	14
T5	20	11.5	9	10.5	-	9
T8	35	7.6	7.4	10.5	7.2	15



[1]Di = Roundabout diameter (m);
[2]We = Width of entry roadway (m);
[3]Wc = Width of circulating roadway (m).

(U-turn, left, through, and right) with vehicle classifications. Digital camera was used to record the data during the a.m. peak period on typical working days (Monday, Tuesday and Wednesday) of April 2010 as Friday and Saturday are the weekend days in Jordan.

5. Analysis and Results

Traffic analysis was conducted in two levels: intersection level and network level.

5.1 Intersection Level Analysis

Although the main purpose of this research was to study alternative traffic options at the network level, an intersection level analysis was conducted first to better understand the existing conditions at each individual intersection, before moving to the network level analysis. The HCS (highway capacity software) was used for the analysis of intersections controlled by traffic signals, and SIDRA software was used for the analysis of intersections controlled by roundabouts. The major scenarios considered in the intersections level analysis included:

(1) Existing condition;

(2) Optimizing the signal timing and phasing;

(3) Optimizing the signal timing and phasing with some geometric improvements (adding free right turns at one approach of intersections T2 and T7, and increasing the roundabout diameter at intersections T4 and T8);

(4) Replacing roundabouts by optimized traffic signals;

(5) Replacing roundabouts by optimized traffic signals with signal coordination;

(6) Using CW (clockwise) one-way traffic ring around Yarmouk University;

(7) Using CCW (counter clockwise) one-way traffic ring around Yarmouk University;

(8) Using two one-way traffic rings in the study network as shown from Fig. 2. King Abdulla II Street was kept running in two ways operation as it is a major entrance to the city and has recently witnessed major traffic improvements.

Fig. 2 Two one-way rings scenarios.

The results of the analysis showed that the last scenario (two one-way rings) was the best alternative based on delay values for intersections (T1, T3, T4, T5 and T7) and the second best alternative for intersections (T2, T6 and T8). The one-way ring scenarios improved the operation at all intersections included in the ring but not necessarily at other intersections in the network. The geometric design improvements scenario was the best for one intersection (T8) and the second best for another intersection (T7). So the one way and/or geometric improvement scenarios were superior to all other scenarios at all intersections.

5.2 Network Level Analysis

The CORSIM was calibrated and validated using the collected field data. It was found that CORSIM may be used for representing the existing conditions with acceptable accuracy (R^2 of 87.6, 89.5, and 96.4 % for traffic speeds, volumes, and maximum queue lengths respectively).

CORSIM produces several MOEs (measures of effectiveness) and on different levels (movement, approach, intersection, link and network). The following are the main MOEs used in this study for comparing different network scenarios:

- Average delay time (seconds per vehicle);
- Average speed (kilometers per hour).

The eight intersection level scenarios were repeated in the network level analysis using CORSIM. Each scenario was compared to the existing condition in terms of percent improvement (increase) in traffic speed as shown in Fig. 3 and improvement (decrease) in traffic delay as shown in Fig. 4.

The best traffic improvements (increase in the average network travel speed and decrease in the average network traffic delay as compared to the existing condition) have been achieved by the two one-way rings, CCW one-way ring, and CW one-way ring scenarios respectively.

It can be seen that the scenario of two rings of one-way streets was the best and provided the highest average network speed (40% increase as compared to the existing condition) and lowest average network traffic delay (47% reduction as compared to the existing condition). This is mainly due to the reduction in the number of signal phases and removal of signals from two T-shaped intersections which ended up having a free traffic operation with only merging and diverging traffic conflicts.

Little benefit was obtained by optimizing the traffic signals (4% speed increase and 7% delay reduction), while adding some geometric improvements provided 20% speed increase and 19% delay reduction. This means that good designed roundabouts with moderate traffic volumes might provide a better network performance than signalized intersections.

Replacing the roundabouts by optimized traffic signals reduced the amount of speed increase to 11% and the amount of delay reduction to only 3%. However, when all signals were coordinated, the speed increase reached 29% and the delay reduction reached 8%. This means that coordinating traffic signals provides more improvement to travel speed than traffic delay.

Using a clockwise one-way ring with optimized and coordinated traffic signals provided 31% speed increase and 19% delay reduction for the network traffic,

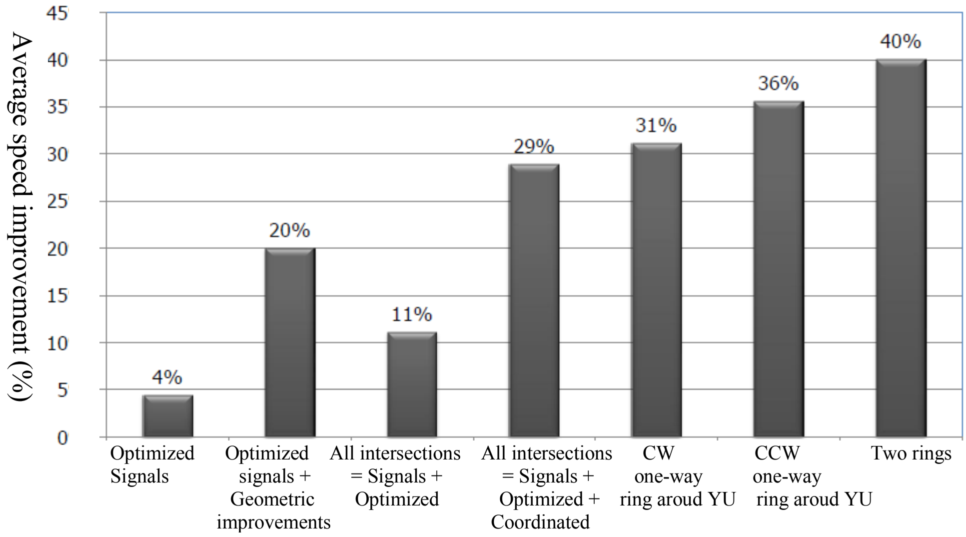


Fig. 3 Percent speed increase for different scenarios as compared to existing condition.

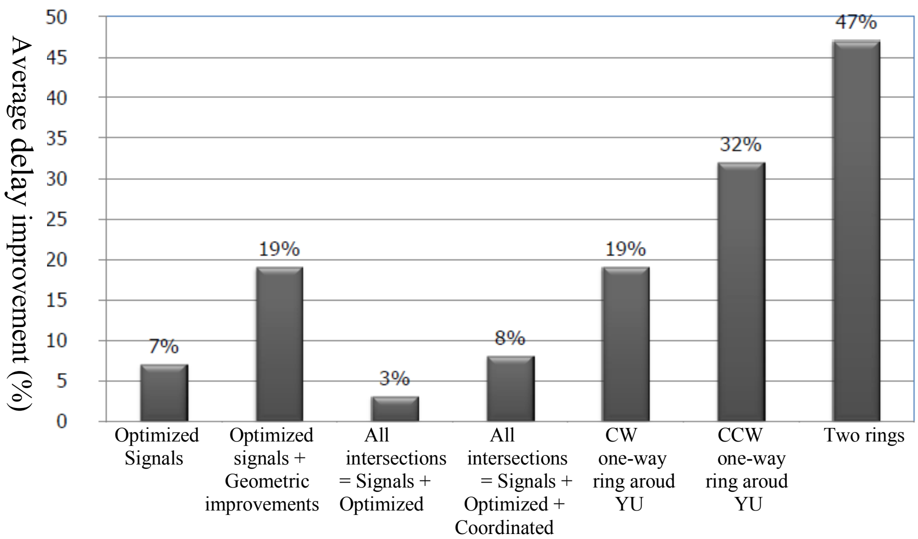


Fig. 4 Percent delay decrease for different scenarios as compared to existing condition.

while using a counter clockwise one-way ring with optimized and coordinated traffic signals provided 36% speed increase and 32% delay reduction, meaning that the counter clockwise is preferred over the clockwise one-way street system. Furthermore, using two rings of one-way streets provided 40% speed increase and 47% delay reduction for the network traffic, meaning that providing more one way-street rings produces higher travel speeds and lower traffic delays over the street network system. The results of this study show that significant operational benefits can be obtained by adopting the one-way traffic system at congested urban street networks.

6. Conclusions

The results of the intersection level analysis based on average intersection traffic control delay showed that the two one-way rings scenario was the best alternative for five intersections and the second best alternative for three intersections. The one-way ring scenarios improved the operation at all intersections included in the rings but not necessarily at other intersections in the study network. The geometric

design improvements scenario was the best for one intersection and the second best for another intersection.

The results of the network level analysis based on the percent increase in the average travel speed and the percent decrease in the average traffic delay as compared to the existing condition showed that the scenario of two rings of one-way streets was the best alternative. It was found that little benefit could be obtained by optimizing the traffic signals, while good designed roundabouts with moderate traffic volumes might provide a better network traffic performance than signalized intersections. Replacing the roundabouts by coordinated and optimized traffic signals provided more improvement to the average sub network travel speed than the average sub network traffic delay. The counter clockwise is preferred over the clockwise one-way street direction.

References

[1] D. Schrank, T. Lomax, The 2003 Urban Mobility Report: Vol. 2, Five Congestion Reduction Strategies and Their Effects on Mobility, Texas Transportation Institute, Texas A&M University, Arlington, Texas, 2003.

[2] M.D. Meyer, A Toolbox for Alleviating Traffic Congestion and Enhancing Mobility, Report Number: IR-054B, Institute of Transportation Engineers, Washington, DC, 1997.

[3] E.D. Arnold Jr, An Evaluation of Congestion-Reducing Measures Used in Virginia. Report Number: FHWA/VA-93-R7. Federal Highway Administration, Washington, DC, 1992.

[4] V. Dixit, Behavioural foundations of two-fluid model for urban traffic. Transportation Research Part C: Emerging Technologies 35 (2013) 115-126.

[5] N. Enustun, Study of the Operational Aspects of One-Way and Two-Way Streets. Report number: Tsd-TR-101 (I)-69, Michigan Department of State Highways, Lansing, Michigan, 1969.

[6] I. Hocherman, J. Bar-Ziv, Safety of one-way urban streets road user behavior: Theory and research, in: the 2nd International Conference on Road Safety, Groningen, Netherlands, 1988, pp. 418-423.

[7] C. Wright, G. Appa, D. Jarrett, Conflict-minimizing traffic patterns: A graph-theoretic approach to efficient traffic circulation in urban areas, Transportation Research Part A: Policy and Practice 23A (2) (1989) 115-127.

[8] J. Wang, L. Yu, F. Qiao, Traffic micro-simulation approach to evaluation of vehicle emissions on one-way versus two-way streets: Case study in downtown Houston, Texas, in: the Transportation Research Board 92nd Annual Meeting, Washington, DC, USA, 2013.

[9] V. Gayah, C.F. Daganzo, Analytical capacity comparison of one-way and two-way signalized street networks, Transportation Research Record: Journal of the Transportation Research Board, 2301(2012) 76–85.

[10] N. Hawkins, H.N. Isebrands, G.L. Fox, Converting CBD roadways from one-way to two-way traffic flow: A case study for the city of Des Moines, Iowa, in: the 2006 ITE Annual Meeting and Exhibit, Milwaukee WI, 2006.

[11] J. Meredith, C. Prem, Central business district traffic circulation study: Kansas City, Missouri, ITE Journal 71 (2) (2001) 26-31.

[12] A. Dbindsa, N.C. Spiller, Traffic Analysis Toolbox Volume X: Localized Bottleneck Congestion Analysis: Focusing on What Analysis Tools Are Available, Necessary, and Productive for Localized Congestion Remediation. Report Number: FHWA-HOP-09-042. Federal Highway Administration, Washington, DC, 2010.

[13] I. Yun, B. Park, Stochastic optimization for coordinated actuated traffic signal systems, Journal of Transportation Engineering 138 (7) (2012) 819-829.

[14] P.S. Lin, S. Rai, E. Gonzalez-Velez, Development of Effective Strategies to Alleviate Traffic Congestion for the Barrier Islands, Final Report, Centre for Urban Transportation Research, University of South Florida, 2007.

[15] N.A. Chaudhary, C.L. Chu, S.R. Sunkari, K.N. Balke, Guidelines for Operating Congested Traffic Signals. Report Number: FHWA/TX-10/0-5998-1, Texas Transportation Institute, Texas A&M University, Texas, 2010.

[16] S. Lin, Z. Zhou, Y. Xi, Analysis of performance criteria for model-based traffic congestion control in urban road networks, in: the Transportation Research Board 92nd Annual Meeting, Washington, DC, USA, 2013.

[17] D. Hale, Incorporation of the HCM2010 trajectory analysis in CORSIM, in: the ITE Technical Conference and Exhibit, San Diego, CA, USA, 2013.

[18] J.E. Hummer, The double-wide design: An inexpensive way to increase the capacity of signalized intersections on four-lane highways, in: the 2nd Urban Street Symposium: Uptown, Downtown, or Small Town: Designing Urban Streets That Work, Anaheim, California, 2003.

[19] K.K. Knapp, K.L. Giese, W. Lee, Urban minor arterial four-lane undivided to three-lane conversion feasibility: An update, in: the 2nd Urban Street Symposium: Uptown, Downtown, or Small Town: Designing Urban Streets That Work, Anaheim, California. 2003.

[20] P. Yu, J A. Meyer, A practical approach for developing a downtown traffic circulation plan–A case study, in: the Transportation Frontiers for the Next Millennium: 69th Annual Meeting of the Institute of Transportation Engineers, Las Vegas, Nevada,1999.

[21] J. Lu, S. Dissanayake, H. Zhou, X. K. Yang, K. Williams, Operational evaluation of right turns followed by U-turns as an alternative to direct left turns. Contract Number: BC-132, Part 3, University of South Florida, Center for Urban Transportation Research, Tampa, FL, 2001.

[22] Traffic Analysis Toolbox Volume IV: Guidelines for Applying CORSIM Micro-simulation Modeling Software, Report Number: FHWA-HOP-07-079. Federal Highway Administration, Washington, DC, 2007.

[23] A. Alomari, Setting Strategies for Traffic Management in Irbid City and its Evaluation Using Simulation. M.S. Thesis, Jordan University of Science and Technology, 2011.

Journal of Control Science and Engineering 1 (2013) 46-54

Long Range Wireless Sensor Network Employing Cooperative Relaying-UAVs via LEO Satellite

David Donald Mrema, Woojin Oh, Jiang Liu and Shigeru Shimamoto

Graduate School of Information and Telecommunication Studies, Waseda University, Tokyo 169-0051, Japan

Abstract: In this study we proposed unique system model for wireless sensor network data telemetry in a wide area and long range. This system is made up with a wide area wireless sensor nodes, UAVs (unmanned aerial vehicles) and LEO (low earth orbit) communication satellite. Wirelessly data signals relaying from sensor nodes on ground to flying UAVs and relaying them to LEO satellite were studied. Three cases of studies were conducted: system without UAV-relay case, one UAV-relay case and two UAV-relay cases. In all the cases, the outage probabilities of data signal channel in Rayleigh fading environment were analyzed and overall system performance was quantified while considering an adaptive decode-and-forward relaying protocol scheme for UAVs. Then, the overall system channel signal-to-noise ratio required in order to achieve a predefined outage probability was analyzed at different channel data rate.

Key words: Unmanned aerial vehicle, low earth orbit satellite, relaying, wireless sensor network.

1. Introduction

The WSN (wireless sensor network) is a network which formed with large number of low power wireless sensor nodes on intended area. WSN has become popular with wide range of applications in a current society; normally, sensor nodes are deployed in means that they can communicate themselves or with the base station wirelessly. The base station is responsible for managing and retrieving data from surrounding nodes. Normally, data from sensor node pass several hopes before reaching base station node. But data relaying through several nodes will degrades the overall system performance in a wide area, since total energy used depends on the number of communication hops. This data relaying will become a trouble on each sensor node due to an exhaustive data processing; mostly this effect will be strong in wide area with randomly deployed sensor nodes.

UAVs (unmanned aerial vehicles) appear to be a promising future intelligent, cheap and efficient device to improve system efficiency by facilitates data collection from wireless nodes in wide area [1-3]. UAVs can act as a central mobile base station or gateway in order to activate sensor nodes which we assumed to be in sleep mode for energy saving then collects data. The UAV requires to having two transceiver payloads for sensor nodes and LEO (low earth orbit) satellite communication in addition with the data processing devices for decode-and-forward the collected data. This collaboration of UAVs and LEO satellite for data relaying from sensor nodes will introduce a lot of benefits in telemetry activities and broaden its applications beyond horizon.

The WSN system with UAV as mobile base station could be simply operated in various kinds of applications as discussed in Refs. [1-3]. However, there are challenges which are brought by UAV to the system in terms of communication to sensor nodes and possibility of real time data retrieval in a wide area. The mobility of UAVs limits the period time for sensor nodes to send the collected data to transceiver

Corresponding author: David Donald Mrema, master, research fields: wireless communication and satellite communication. E-mail: donsonster@gmail.com.

payload on UAV, also it influences high data packet error rate. Also, the ability of UAV to relay the collected data to mission center is limited with UAV's transceiver communication range. However, Ho et. al [1] proposed a MAC (media access control) protocol for the system in data collection section between UAV and sensor nodes that are based on both TDMA (time divisions multiple access) schemes and FRA (frame based random access) scheme. The randomly deployed sensor nodes were grouped based on level of receiving sensor activation signal power from UAV, the signal power level varies from low at the edge of UAV ground coverage area to high at center. When UAV is in motion, the major groups are divided into sub-groups (sensors "in front of" and "behind" UAV direction), so as to give higher priority to sensors with lesser remaining connection time with UAV. Then, data transmission carried out in TDMA scheme mode when the channels collide freely due to involvement of priority of data transmission within sensor node sub-groups. The data relaying mechanism from UAV to mission center was not discussed in this study.

Our objectives are different from the above study. We upgrade the system to have features of data relaying in a long range also near-real time. Most system for WSN employing UAV discussed in Refs.[1-3] normally store the collected data in a storage payload device on UAV or relay them to a nearby base station in a mobile terrestrial network. In study, we need a system without depending on terrestrial networks, which can survive in any disaster, or is applied anywhere- regardless of terrestrial network coverage.

Since LEO communication satellites network is the promising future in telecommunications services in a wide area compared to current terrestrial networks [4], we suggest the data relaying system model that can combine both UAVs and LEO satellite networks. The introduction of UAVs in LEO satellite network for data relaying applications will increase the communication range and near-real time data transmission in telemetry activities.

This synergy system model can be used for disaster management (i.e., tsunami monitoring) with WSN, as seen in Fig. 1. In this detailed system model both UAV control signal from mission center and intended mission data from sensor nodes are relayed from UAV via LEO satellite. The overall data collections start from randomly or organized deployed sensor nodes, which is sent to multiple UAVs flies above the disaster area and then relayed them to far-end mission center via LEO communication satellites networks.

Introducing relaying channel between UAVs and LEO satellite was necessary so as to upgrade UAVs communication range. Besides the necessity of introducing relaying UAVs, the relaying strategy is of great importance, there are major three relaying strategies: multi-hop, multi-route and request relaying networks [5-7]. In our study, we focused on multi-route relaying networks, with respect to a predefined transmission scheme, where the source transmits information to several relay, simultaneously those relays retransmit information either, or both at the same time. Hence, the destination receives information from several relays and direct paths from the source.

Furthermore, we evaluate the system performance by finding the outage probability with zero, one and two UAVs as a relay with decode-and-forward transmission protocol with repetition coding scheme (means each relays transmit new version of source signal by applying the same kind of coding). The channels were regarded as independent but not identically distributed, and the selection combining is used at the destination (LEO satellite) and the moment generation function was also not used in deriving the system performance general expression. The rest of this paper is organized as follows. Section 2 discusses the network model and channel access scheme between UAVs to ground and UAVs to LEO satellite. Section 3 outage probability is presented and Section 4 describes simulation results. Section 5 gives conclusion and future works.

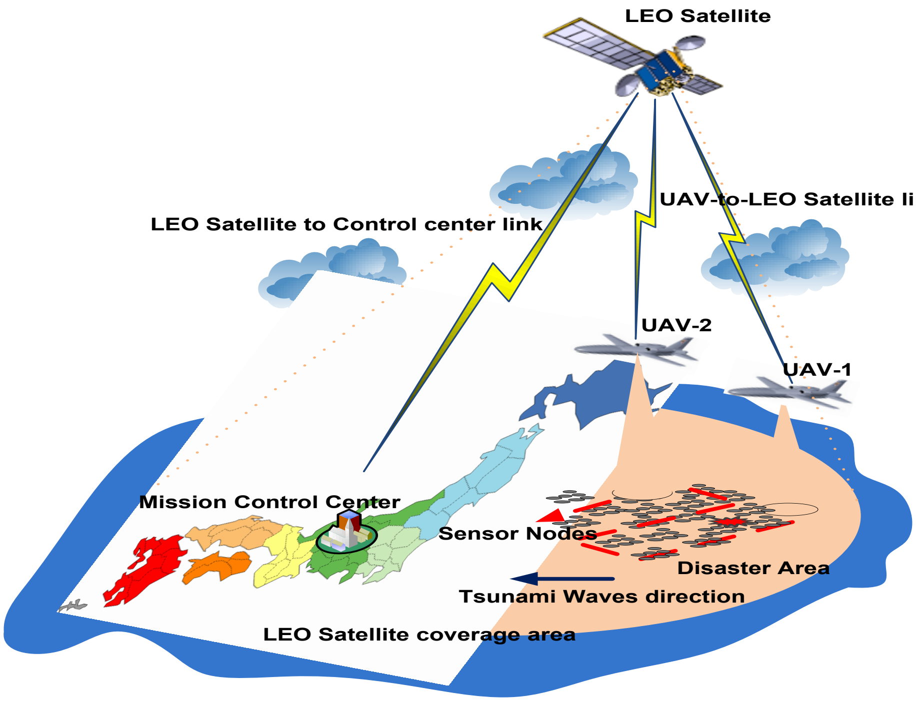


Fig. 1 Layout description of tsunami monitoring system model with WSN-UAV and LEO satellite.

2. Network Model and Channel Access

2.1 Network Model

The communication system illustrated in Fig. 1 can be expressed in a model form as Fig. 2 shows, where we combine sensor nodes as one source S, LEO satellite networks as one destination D and N relay UAVs as (R_1, R_2, …, R_N), i.e., there are N + 1 transmitting terminals (source and relay). The received signal $y_n(t)$ at destination for each relay n is given by:

$$y_n(t) = h_{sn} x_{sn}(t) + z_n(t) \tag{1}$$

where, $n \in [1, 2, \ldots, N]$ represent the number of the relay UAVs and subscript s stands for source, h_{sn} is the channel coefficient between source (sensor nodes) and relays (UAVs) and $Z_n(t)$ is AWGN (additive white Gaussian noise) with one sided spectral density N_0. The received signal $y_d(t)$ at the destination (LEO satellite) can be expressed as:

$$y_d(t) = h_{sd} x_{sd}(t) + \sum_{n=1}^{N} h_{nd} x_{nd}(t) + z_d(t) \tag{2}$$

where, the first term is the direct path from source to destination, the subscript d stands for the destination, h_{sd} represents channel coefficient for n relays to destination and $Z_d(t)$ is AWGN. Since the relay performs the decode-and-forward of the received signals from the source, the transmit signal of the relay is the estimate version of the received source signal, i.e., $x_{nd}(t) = \hat{x}_{sn}(t)$. The channel coefficient h_{sn} and h_{nd} are modeled as independent circular-symmetric complex valued Gaussian random variables. We model the channel as Rayleigh fading, i.e., the magnitude $|h_{sn}|$ and $|h_{nd}|$ channel coefficients follow a Rayleigh distribution. Therefore, the channel power $|h_{sn}|^2$ and $|h_{nd}|^2$ are exponentially distributed with mean values $\sigma_{sn}{}^2$ and $\sigma_{nd}{}^2$ which can be given by the distance-dependent path losses. The block fading state was considered where the channel influences are constant over one transmission frame of duration T.

2.2 Channel Access

Channel access is a major criterion for the performance of radio systems in wireless communications.

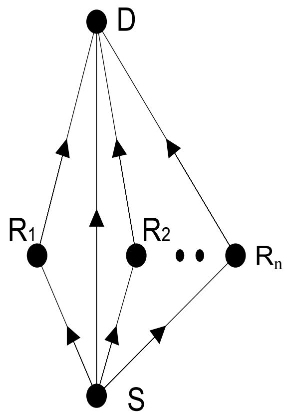


Fig. 2 The general vector representation of data signal channels from source and *n* relays to destination.

We divide the discussion of channel access in two groups: UAVs to ground communication where UAV communicates with sensor nodes and UAVs as relay to LEO satellite networks.

2.2.1 UAVs to Ground Communication Channel

In UAVs to ground channel access, first UAV sends the activation signal on the coverage area in which sleeping sensor nodes are randomly deployed, if the activation signal is less than threshold value then sensor nodes will stay in sleep mode. This was considered in order to give sensor nodes longer life time by reducing its retransmission effect due to poor channel condition. The activated sensor nodes will send their data periodically to UAV as many as possible during the availability of activation signal. The sensor nodes within UAV coverage area are divided to sub-groups so as to allow time interval in channel access for each active sensor nodes. The priority was set within sub-groups; higher priority was set to those groups which are behind UAV during its motion, since they have shorter time duration in channel access and data transmission. Lower priority was set to those groups which are near and in-front of UAV direction motion. Since sensor nodes on ground are not connected to each other and their position is undefined proper channel access must be considered. Time division multiple access schemes was selected to carry out commands for channel access between UAV and sensor nodes [1].

2.2.2 UAVs to LEO Satellite Communication Channel

The channel access between UAVs to LEO satellite is assumed to be carried out under time division multiple access schemes, where n relays (UAVs) transmit the collected information signals from source to a frame with duration T. The frame duration T is divided in $N + 1$ sub-frame of length $T/ (N + 1)$ and at each sub-frame a transmission rate of $(N + 1) R$ was assumed, where R is channel data rate [5, 6]. The overall system energy is regarded to be constant for fair comparison. The energy of direct transmission channel and that of source via relays (UAVs) with source transmit power P must be equal to the energy from source and n relays transmit to $N + 1$ sub-frames of duration $T/ (N + 1)$. Each terminal transmits with power P, then the overall system energy becomes PT.

Fig. 3 gives detailed information for data transmission from source to destination via relays in sub-frames bases for both one-relay and two-relay case. Starting with one-relay case, source transmits to both relay and destination with power P in duration of T/2. Then relay decode received information signal from source and then transmits it to destination with transmit power P in remained duration of T/2 in total sub-frame. Then as for two-relay case, source transmit to both relays and destination with power P and duration of T/2, then those two-relay transmits received information signals and sent to destination concurrently, each with power P/2 within sub-frame of duration T/2. For both one- relay and two-relay case, the data rate is twice when comparing to the system with direct transmission without relay. However, in two-relay case the inter-relay interference will occur since there is no transmit agreement between them, but this effect can be taken care by introducing orthogonal data transmitting between relays [8].

3. Outage Probability

3.1 General Expression for Independent Path

The different path signals of information from

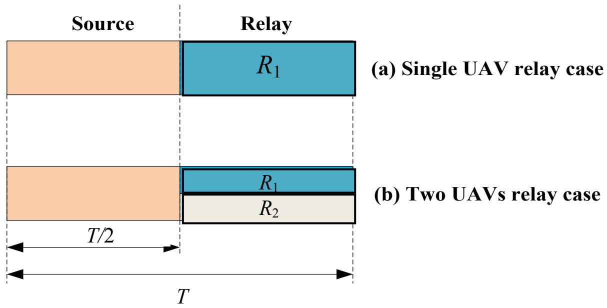


Fig. 3 Sub-frames allocations for source and relays data transmission to destination in (a) one-relay and (b)two-relay cases.

source to relays and from relay to destination are regarded as statistically independent. Since the selection combining is applied at the destination then, for the relay system to fail there must be a failure of all connections from source to destination and source to destination via relays. This can happen either from source to relay failure or after the relay decode received source information signal and fail to transmit it to destination point. The overall system performance will be analyzed based on these combinatorial failure events for data transmission from source to destination via relays.

Let us define the outage (fail) event A_{ij} = {transmission from terminal i to terminal j failed}. Then the complement of failure event A_{ij} is represented as $A^c{}_{ij}$. The mutual probability of failure for N relays is given by:

$$\sum_{1\le i_1<...<i_n\le N} \Pr\left(A_{si_1},......A_{si_n}\right) = \sum_{1\le i_1<...<i_n\le N} \Pr\left(\bigcap_{m=1}^{n} A_{si_n}\right) \quad (3)$$

where, $1 \le i_1 \le \cdots \le 1_n \le N$ represent all $\binom{N}{n}$ combinations of failures. Also, when N-n relays fail to send information successful to destination after decode it, and source to destination failure was considered. Then overall data transmission outage probability P_{out} can be expressed as:

$$P_{out} = \sum_{n=0}^{N} \sum_{1\le i_1<...<i_n\le N} \Pr\left(A_{sd} \bigcap_{m=1}^{n} A_{si_m} \bigcap_{\substack{q=1 \\ q\neq\{i_1..i_n\}}}^{N} A^c{}_{sq} A_{qd} \right) \quad (4)$$

Where, A_{sd} is the source to destination outage event. Each information signal path is statistically independent, Eq. (4) can be written as:

$$P_{out} = \Pr(A_{sd}) \sum_{n=0}^{N} \sum_{1\le i_1<....<i_n\le N} \prod_{m=1}^{n} \Pr(A_{si_m}) \prod_{\substack{q=1 \\ q\notin\{i_1..i_n\}}}^{N} \Pr(A^c{}_{sq}) \Pr(A_{qd}) \quad (5)$$

The outage probability of the Rayleigh fading environment channel with average $\bar{\gamma}_{ij}$ signal-to-noise ratio at terminal j and γ_{th} threshold signal-to-noise ratio required for reliable communication is [5, 6]:

$$\Pr(\gamma_{ij} \le \gamma_{th}) = \Pr(A_{ij}) = 1 - \exp\left(-\frac{\gamma_{th}}{\bar{\gamma}_{ij}}\right) \quad (6)$$

The threshold value γ_{th} depends on data rate R. Note that, $\bar{\gamma}_{ij} = E\{|h_{ij}|^2\}\gamma_{ij} = \sigma^2{}_{ij}\gamma_{ij}$, where $\sigma^2{}_{ij}$ is the mean value of the exponential random variable of channel coefficient $|h_{ij}|^2$ and $\gamma_{ij} = P_{ij}/N_0$ is the signal-to-noise ratio without fading. When substituting Eq. (6) into Eq. (5), the outage probability can be expressed as:

$$P_{out} = \left(1 - \exp\left(-\frac{\gamma_{th}}{\bar{\gamma}_{sd}}\right)\right) \sum_{n=0}^{N} \sum_{1\le i_1<....<i_n\le N} \prod_{m=1}^{n} \left(1 - \exp\left(-\frac{\gamma_{th}}{\bar{\gamma}_{si_m}}\right)\right) \prod_{\substack{q=1 \\ q\notin\{i_1..i_n\}}}^{N} \exp\left(-\frac{\gamma_{th}}{\bar{\gamma}_{sq}}\right)\left(1 - \exp\left(-\frac{\gamma_{th}}{\bar{\gamma}_{qd}}\right)\right) \quad (7)$$

The analysis of overall system performance in terms of outage probability for one-relay and two-relays UAVs relay cases will follow in the next subsections. These analyses follow the general expressions and assumptions made in Refs. [5, 6]. Those conditions are:

(1) If the transmission from source to relay fails (P_{sr}), then the transmission from source to destinations fails with probability of 1 (P_{sd} = 1);

(2) Each channel path from source to destination and relay to destination are independent.

3.2 The One UAV-Relay Case

The system outage probability with one UAV-relay

case can be derived from the selection combining of signal path from source to relay, source to destination and relay to destination. Refering to Eqs.(1) and (2) , the signals at relay and destination can be represented as:

$$\begin{aligned} y_r(t) &= h_{sr}x_{sr}(t) + z_r(t) \\ y_d(t) &= h_{sd}x_{sd}(t) + h_{rd}x_{rd}(t) + z_d(t) \end{aligned} \tag{8}$$

Where, $x_{sr}(t)$ = $x_{sd}(t)$ = $x(t)$ for $t \in [0;T/2)$ and $x_{rd}(t) = \hat{x}(t - T/2)$ for $t \in [T/2;T)$.

The two step of our transmission scheme i.e. source and relay both transmit as describe in Fig. 3(a). The end-to-end data rate for each path is given by $(N+1)$ R [6].

$$\gamma_{th} = 2^{2R} - 1 \tag{9}$$

$$f(R,\gamma) \Leftrightarrow \frac{\gamma_{th}}{\gamma} = \frac{2^{2R}-1}{\gamma} \tag{10}$$

Let $f(R,\gamma)$ represent the ratio of individual channel path signal-to-noise ratio to threshold value. From Eq. (5), the outage probability for one UAV relay case is $P_{out(1)} = P_{sd}P_{sr} + P_{sd}(1-P_{sr})P_{rd}$, then by using Eq. (10) to do some mathematical manipulation, then;

$$\begin{aligned} P_{out(1)} = &\left(1 - \exp\left(-\frac{f(R,\gamma)}{\sigma^2{}_{sd}}\right)\right) \\ &\left(1 - \exp\left(-f(R,\gamma)\left(\frac{1}{\sigma^2{}_{sr}} + \frac{1}{\sigma^2{}_{rd}}\right)\right)\right) \end{aligned} \tag{11}$$

3.3 The Two UAVs-Relay Case

In the two UAVs-relay cases we assume that relays transmit simultaneously during the relay sub-frame transmission time as shown in Fig. 3b with interference-free between each other. The signal received at relays points are:

$$\begin{aligned} y_{r1}(t) &= h_{sr1}x_{sr1}(t) + z_{r1}(t) \\ y_{r2}(t) &= h_{sr2}x_{sr2}(t) + z_{r2}(t) \end{aligned} \tag{12}$$

While the received signal at destination becomes:

$$\begin{aligned} &y_d(t) = \\ &h_{sd}x_{sd}(t) + h_{r1d}x_{r1d}(t) + h_{r2d}x_{r2d}(t) + z_d(t) \end{aligned} \tag{13}$$

Then, information signal from source $x(t)=x_{sr1}(t)=x_{sr2}(t)=x_{sd}(t)$ for $t \in [0;T/2)$ and $x_{r1d}(t) = x_{r2d}(t) = \hat{x}(t - T/2)$ for $t \in [T/2;T)$. The conditions for both relays to have a fair comparison are transmit with equal power P/2. The threshold signal-to-noise ratio is the same as in Eq. (8) which can be derived from simultaneously transmission protocols of two UAV-relays on the same sub-frame. Then, the ratio between threshold and instantaneous signal-to-noise ratio becomes:

$$\widetilde{h}(R,\gamma) \Leftrightarrow \frac{\gamma_{th}}{\gamma} = \frac{2^{2R}-1}{\gamma} \tag{14}$$

Also, from Eq. (5) the outage probability for two UAVs-relay cases when both transmit simultaneously can be computed as

$$\begin{aligned} &P_{out(2)} = \\ &P_{sd}(P_{sr1}P_{sr2} + (1-P_{sr1})P_{sr2}P_{r1d} + P_{sr1}(1-P_{sr2})P_{r2d} \\ &\quad + (1-P_{sr1})(1-P_{sr2})P_{r1d}P_{r2d}) \end{aligned} \tag{15}$$

Also Eq. (15) can be expressed as Eq. (11) by substituting Eq. (14) into Eq. (15)

4. Simulation Results and Discussions

In this section, we give out the simulation results of our study by using MATLAB, computer programming software.

4.1 Outage Probability

The channel path was derived from simple model where the path loss between nodes *i* and *j* depends on distance d_{ij} and path loss exponent *k*, i.e., $\sigma^2{}_{ij} \; \alpha \; d^{-k}{}_{ij}$. Normally, the value of path loss exponent ranges from 2 to 4. In our analysis we use $k = 2$ for channel path between sensor nodes (source) and UAVs (relays) since we assume free space propagation with line of sight. While we use $k = 4$ for channel path between UAVs (relays) and LEO satellite (destination). All channel path distances were normalized with respect to source-destination distance, i.e., $d_{sd} = 1$. For that case $\sigma^2{}_{sd} = 1, \sigma^2{}_{sr_i} = d^{-k}{}_{sr_i}$ and $\sigma^2{}_{r_id} = d^{-k}{}_{r_id}$ where $i = 1, 2$.

The result as shown in Fig. 4 shows the system outage probability at different signal-to-noise values and data rate R = 1.5 bit/s/Hz. We take the position of relays to be $d_{sr1} = d_{r2}$ = 0.00125, which reflects a normalized value of 500 m from LEO satellite (destination) with orbital distance of 400 km from earth ground level. Results show that system with relay outperforms the one without relay from SNR (signal to noise ratio) > 14dB throughout. The higher diversity order of 2 and 3 for one-relay case and two-relay case respectively was the main reason for better performance compared to 1 diversity order of system channel without relay [9]. Note that, system channel Doppler shift effect was assumed to be taken care by receiver with either frequency stabilization, re-tuning mechanism to adopt the changes over time or receiver with beam forming antenna [10].

4.2 Required SNR at Predefined System Outage Probability

The analysis of system Signal-to-Noise Ratio (SNR) required to achieve a predefined outage probability as function of data rate was aiming to understand the amount of energy the system must use to attain a specific data channel transmission qualities. Figs. 5 and 6 were obtained from the derivation of Eqs. (11) and (15) with mathematical approximation i.e., $1-\exp(-A) \approx A$. Fig. 5 shows that system data transmission channel without relay performs worst compared to other two with relay case for value of R < 5 bits/s/Hz and outage probability of $P_{out} = 10^{-3}$. When the system data rate is less than 2 bits/s/Hz, it shows that with two-relay case we can save approximately 44% of overall system energy required to be used in data transmission when compared to system without relay. The system requires high SNR at high data rate for both cases without relay and with relay, but one-relay case and two-relay case graphs give the same slope for the entire range of system data rate. This behavior results from the amount of time slots allocated for the source transmission as a ratio to the total frame time [8].

In Fig. 6, we can observe the reliance of SNR values on P_{out} within the same range of system data rate. We reduce the outage probability by 10^{-1} factor to have $P_{out} = 10^{-4}$. The results show that system without relay increases its SNR value by 10 dB. The one-relay case increases by 5 dB and two-relay case increases by 3.24 dB (comparing Figs.5 and 6). Also, we tried to compare the amount of energy used with new value of outage probability. In two-relay case, we can save approximately 50% of system energy used for data transmission with data rate less than 2

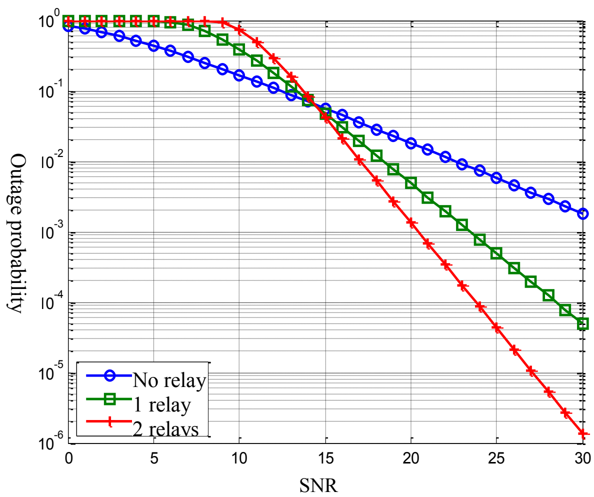


Fig. 4 The overall system data transmission outage probability for without relay case, one relay case and two relays case against channel signal-to-noise ratio.

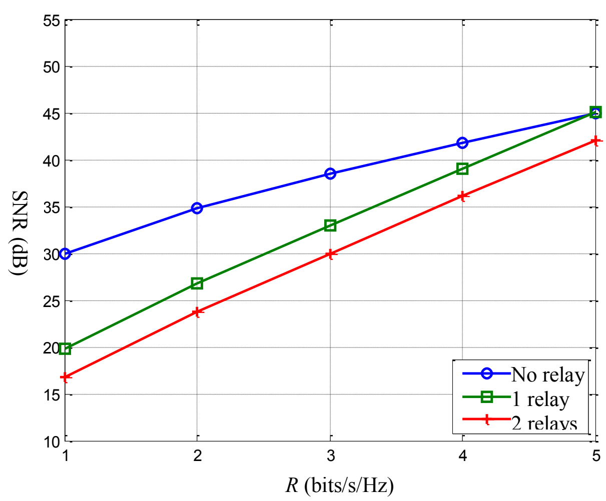


Fig. 5 Signal-to-noise ratio (dB) against system channel data rate R (bits/s/Hz) at predefined system outage probability set to Pout = 10^{-3}.

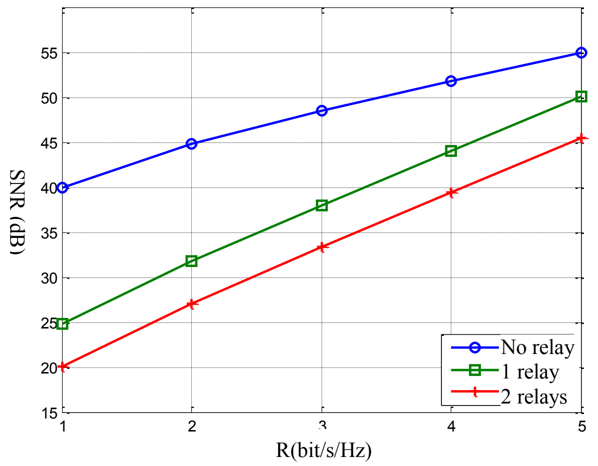


Fig. 6 Signal-to-noise ratio (dB) against system channel data rate *R* (bits/s/Hz) at predefined system outage probability set to Pout = 10^{-4}.

bit/s/Hz when compared to system without relay case. (Note; SNR value is proportional to system energy used for data transmission). This shows that system with relay requires less SNR values to achieve better performance with respect to outage probability and increase system life time due to less energy consumption.

5. Conclusion and Future Works

In this study we have proposed a near real time and long range wireless sensor network employing cooperative UAV-relays via LEO satellite. The analysis of overall system data transmission from ground deployed sensor nodes (source) to LEO satellite (destination) via UAV (relay) was carried out based on three cases: data transmission from source to destination (1) without relay case, (2) one-relay case and (3) two-relay case. The outage probability expression was used to quantify the system performance in all three cases. The outage probability expression was derived based on independent but not identically distributed propagation paths of data transmission channels within a decode-and-forward relaying transmission protocol. In channel access, both source and relay have a specific individual sub-frame within main frame, and for the two-relay case both relays transmit simultaneously on the same sub-frame with half the power to ensure fair comparison to system without relay case.

The system with relay shows better performance when compared to the case without relay. This is due to relaying strategies achieve full order of diversity, which leads to lower value of system outage probability.

The SNR required for a specific data rate with predefined system outage probability level was analyzed. The two-relay case shows the best performance compared to other cases since it needs less SNR to attain specific outage probability. Also, we have realized that the change of 10^{-1} of outage probability results to an increase of SNR value by 10 dB for system without relay case, 5 dB for one-relay case and 3.24 dB for two-relay case.

In future studies we plan to continue with this study so as to analyze the system performance with different positions of relay-UAVs, also the optimal number of relay-UAVs. Furthermore, we will study the inter-satellite communication in LEO satellite networks in order to figure out the overall system data transmission up to mission control center.

Acknowledgments

The authors would like to thank Japanese Government Scholarship through MEXT (Ministry of Education, Culture, Sports, Science and Technology), and Grant-in-Aid for Scientific Research© No. 23560473 for their fund support during this study.

References

[1] D. Ho, J. Park, S. Shimamoto, Performance evaluation of the PFSC based MAC protocol for WSN employing UAV in racian fading, in: IEEE WCNC, 2011.

[2] Civilian applications of unamaaned aircraft systems, NLR Air Transport Safety Institute Newsletter [Online], http://www.nlr-atsi.nl/id-14865.html (accessed January 18, 2013).

[3] R. Jafari, A. Encarnacao, A. Zahoory, F. Dahiri, H. Noshadi, M. Sarrafzadeh, Wireless sensor networks for health monotoring, in: Proc. Mobile and Ubiquitous Systems: Networking and Services, July 2005.

[4] L.J. Ippolito, Satellite Communications Systems Engineering, John Wiley & Sons Ltd., 2008.

[5] P. Herhold, E. Zimmermann, G. Fettweis, Cooperative multi-hop transmission in wireless netwroks, Journal on Computer Networks 49 (3) (2005) 299-324.

[6] T. Renk, H. Jaekel, C. Kloeck, F.K. Jondral, General expression of outage probability in cooperation networks, European Transaction on Telecommunications 20 (6) (2009) 537-548.

[7] O. Mazen, A. Mohamed-Slim, End-to-end performance of transmission system with telays over Rayleigh-fading channels, in: IEEE VTC (Vehicular Technology Conference), Sept., 2002.

[8] B. Rankov, A. Wittneben, Spectral efficienct signaling for half-duplex relay channels, in: Conference Record of the 39th Asilomar Conference on Signal, Systems and Computers, Pacific Grove CA, USA, Oct. 28-Nov.1, 2005.

[9] A. Almouti, A simple transmit diversity technique for wireless communication, IEEE Journal on Selcted Area in Communication 16 (8) (1998) 1451-1458.

[10] Iskandar, S. Shimamoto, Channel characterization and performance evaluation of mobile communication employing stratospheric platform, IEICE Trans. on Commun E89-B (3) (2006) 937-944.

Journal of Control Science and Engineering 1 (2013) 55-59

Electromagnetic Coupling Modeling of Microstrip Lines Used in Switching Power Supplies

Nassima Tidjani[1], Jean-Charles Le Bunetel[2], Ali Ouchar[1] and Yves Raingeaud[2]

1. LTSS (Laboratory of Telecommunication Signals and System), University of Amar Telidji, Laghouat 03000, Algeria

2. GREMAN (Research Group on Materials, Microelectronics, and Acoustics Nanotechnology), University of Tours, Tours 37200, France

Abstract: One of the major problems in electromagnetic compatibility is the interference due to crosstalk on the PCB (printed circuit board) in the SiP (system-in-package) technology. In this work, electromagnetic coupling modeling of coupled microstrip lines used in switching power supplies is presented. The different blocks which are connected through buses realized on PCBs will be considered as a coupled transmission line. The coupled microstrip transmission lines are modeled by a mathematical model based on the MoM (method of moments). This model is done by Matlab routines. A SPICE (simulation program with integrated circuit emphasis) model is proposed for the prediction of the NEXT (near-crosstalk) and the FEXT (far-crosstalk) at the terminals of the command line. This command line is often victim of electromagnetic interference emitted by a transmission line source close. Some solutions against crosstalk depending on the physical parameters of the coupled transmission lines are proposed.

Key words: Crosstalk, PCB, SiP, microstrip, NEXT, FEXT.

1. Introduction

To make increasingly small PCB (printed circuit board) boards and meet the needs of today's market, electronic components are increasingly miniaturized and close to each other. The SiP (system-in-package) technology aims to integrate several integrity circuits, passive components, connectors and antennas in one box to create fully functional electronic subsystems [1]. But the miniaturization causes the development of electromagnetic interference like crosstalk between the various blocks constituting a PCB board [2, 3]. In this paper, electromagnetic coupling modeling of coupled microstrip lines used in switching power supplies is presented. These buses realized on PCBs will be considered as coupled transmission lines. The study of crosstalk requires knowledge of per unit length requires, knowledge of per unit length parameters and coupling coefficients between coupled transmission lines. In order to model the electromagnetic coupling present in PCB board, a SPICE (simulation program with integrated circuit emphasis) model is proposed. The model is based on the parameters per unit length of coupled transmission lines extracted by the method of moments. This model allows the prediction of the NEXT (near-crosstalk) and the FEXT (far-crosstalk) between coupled transmission lines routing the control signal used in switching power supplies. From the study of the proposed SPICE model, some solutions against crosstalk depending on the physical parameters of the coupled transmission lines are proposed.

2. Coupling Coefficients Modeling

The cohabitation of different blocks on the same PCB board increases problems, due to different electromagnetic couplings as crosstalk [4].

Corresponding author: Jean-Charles Le Bunetel, assistant professor, research fields: electromagnetic compatibility, power electronics, transmission lines, communication power line. E-mail: lebunetel@univ-tours.fr.

The origin of the electromagnetic couplings parasites is linked to the internal activity of the components. The circulations of current generate a magnetic field, while the fast switching of signal creates an electric field [5].

The crosstalk prediction is based on knowledge of the coupling coefficients between the coupled transmission lines. There are many methods in the literature for the extraction of the coupling coefficients [6].

In this paper, we used a mathematical model developed in Refs. [7-11], which based on the method of moments.

This mathematical model allows the extraction of parameters per unit length of coupled transmission lines. A Matlab routine was developed for the determination of per unit length elements in function of the physical parameters (spacing, strip width, substrate).

3. SPICE Model

Telegraph equations governing the coupled transmission lines illustrated in Fig. 1 are written as follow:

$$\frac{\partial v_1(z,t)}{\partial z} + L_{11}\cdot\frac{\partial i_1(z,t)}{\partial t} = -L_{12}\cdot\frac{\partial i_2(z,t)}{\partial t} \quad (1)$$

$$\frac{\partial i_1(z,t)}{\partial z} + C_{11}\cdot\frac{\partial v_1(z,t)}{\partial t} = -C_{12}\cdot\frac{\partial v_2(z,t)}{\partial t} \quad (2)$$

$$\frac{\partial v_2(z,t)}{\partial z} + L_{22}\cdot\frac{\partial i_2(z,t)}{\partial t} = -L_{21}\cdot\frac{\partial i_1(z,t)}{\partial t} \quad (3)$$

$$\frac{\partial i_2(z,t)}{\partial z} + C_{22}\cdot\frac{\partial v_2(z,t)}{\partial t} = C_{21}\cdot\frac{\partial v_1(z,t)}{\partial t} \quad (4)$$

The SPICE program automatically determines and solves thee quations depending on the frequency. We have $L_{11} = L_{22}$, $C_{11} = C_{22}$, $C_{12} = C_{21}$, $L_{12} = L_{21}$, because the coupled lines are symmetric.

The per unit length parameters of coupled microstrip lines based on the method of moments are introduced into a SPICE model illustrated in Fig. 2. The magnetic and electrical couplings are realized through the mutual inductance L_{12} and mutual capacity C_{12}, respectively.

This model will be used to study the crosstalk by calculating the NEXT and the FEXT in function of physical parameters, and also the rise time of the input signal. This study will allow us to find the optimal distance of tolerance, which allows limiting the effects of the crosstalk. To validate the electrical model, experimental measurements of the NEXT and the FEXT on a PCB card were compared with simulation results of the electric model. The PCB card used consists of two microstrip lines having a width W = 3 mm, a space S = 8 mm, height of the substrate H = 1.5 mm, and a length l = 25 cm, and a ground plane on the other face, and we use four connectors, port 1 and port 2 for the aggressor line, and port 3 and port 4 for the victim line. The input signal is a square wave signal of 4 V amplitude, and frequency 1 MHz, with a rise time t_r = 10 ns applied to port 1. Port 2 is loaded with a matched load.

The measurements of the NEXT and the FEXT are performed on port 3 and port 4. Figs. 3 and 4 represent the NEXT and the FEXT collected from the SPICE model and from the measurement. We observe a good reproduction of the electromagnetic coupling phenomenon of the electric model.

Fig. 1 Photography of PCB.

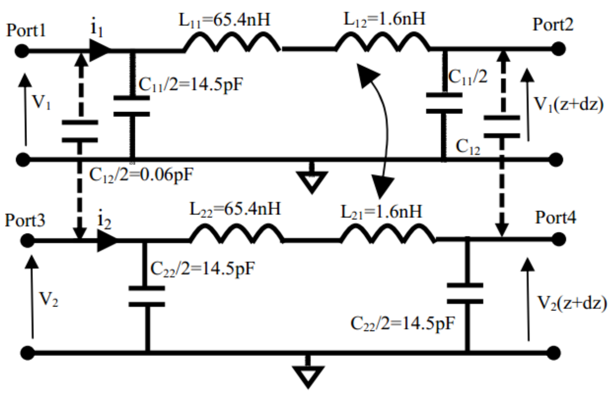


Fig. 2 SPICE model of coupled microstrip lines.

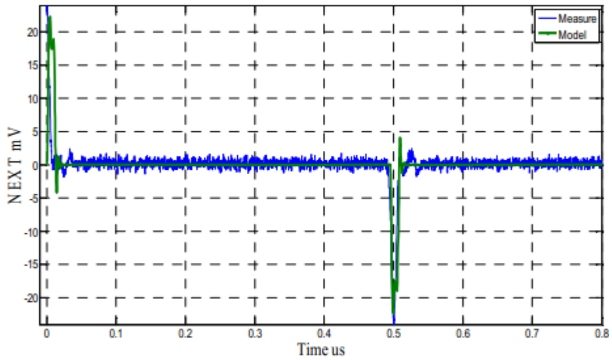


Fig. 3 The NEXT by the model and the measurement.

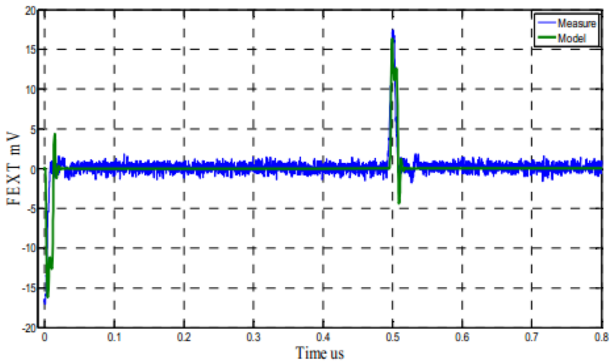


Fig. 4 The FEXT by the model and the measurement.

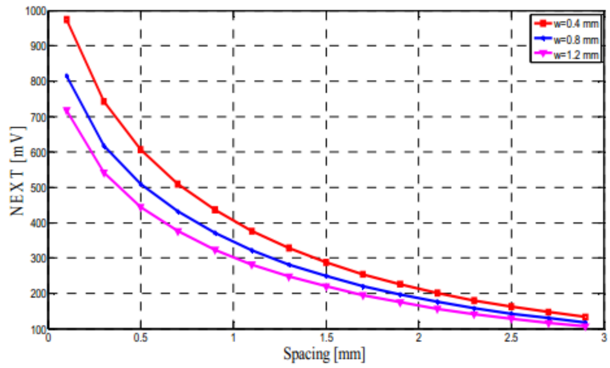


Fig. 5 The NEXT for different strip widths.

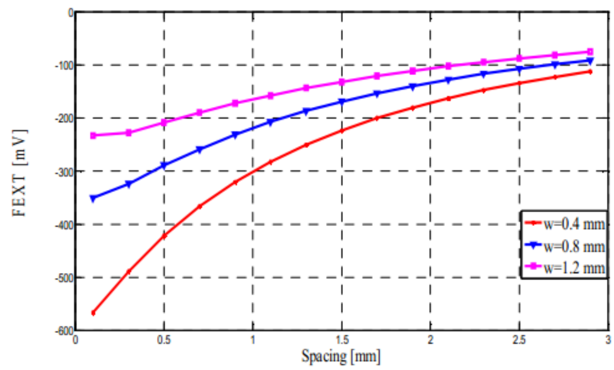


Fig. 6 The FEXT for different strip widths

4. Crosstalk Solutions

The proposed model allows us to study the crosstalk by proposing abacus for helping the designers of PCBs. This model predicts crosstalk in function of physical parameters of coupled transmission lines and the input signal dynamic.

4.1 Influence of Strip Width

As a function of the distance between two coupled microstrip, we studied the influence of the strip width W on the crosstalk. The calculation of per unit length parameters was performed for different spacing values between the coupled microstrip and for three strip widths: W_1 = 0.4 mm, W_2 = 0.8 mm and W_3 = 1.2 mm.

Figs. 5 and 6 show respectively the NEXT and the FEXT depending on the spacing between the coupled microstrip lines, for three fixed strip widths. We observe that the NEXT and the FEXT decreases exponentially with the increase of the spacing between the coupled transmission lines. In addition increasing track width reduces the crosstalk.

4.2 Influence of Strip Line Length

To study the contribution of the length of coupled lines on the crosstalk, we compute the parameters per unit length by setting the strip width at W = 1 mm, and for three lengths of line: l_1 = 4 cm, l_2 = 16 cm, and l_3 = 60 cm.

Figs. 7 and 8 represent the voltages of the NEXT and the FEXT available for different lengths. It is observed that the influence of the length of lines is non-linear, and it is more important for the NEXT than for the FEXT. The crosstalk decreases as the length of coupled lines decreases.

4.3 Influence of Dielectric Permittivity

Among the physical characteristics, we have the dielectric permittivity of the substrate used. To see the impact of a high or low dielectric permittivity on the crosstalk we chose three cases of dielectric permittivity: ε_{r1} = 2.2, ε_{r2} = 4.7, and ε_{r3} = 12.

Figs. 9 and 10 illustrate the variation of the NEXT and the FEXT for different dielectric permittivity. We

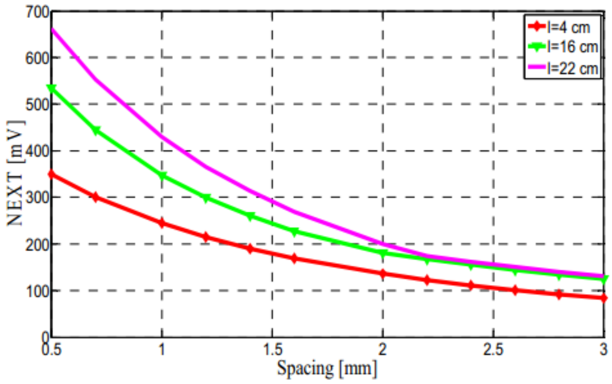


Fig. 7 The NEXT for different lengths.

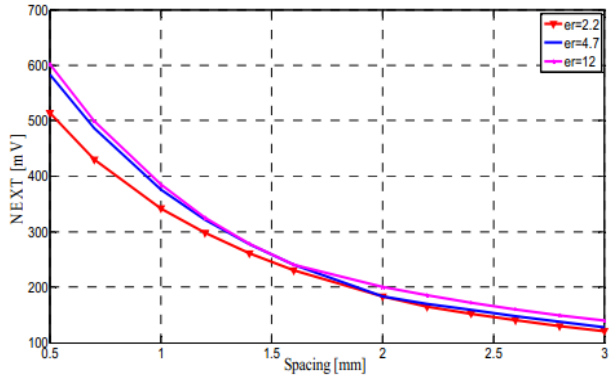


Fig. 9 The NEXT for different dielectric permittivity.

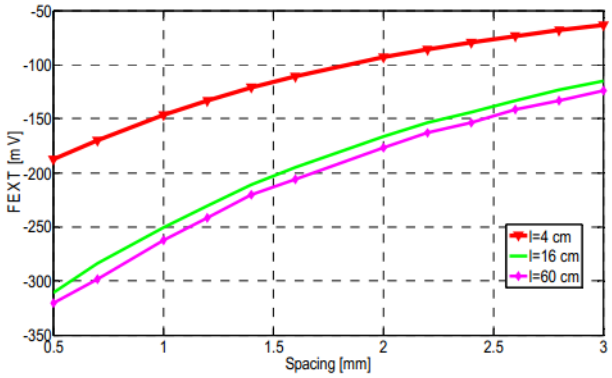


Fig. 8 The FEXT for different lengths.

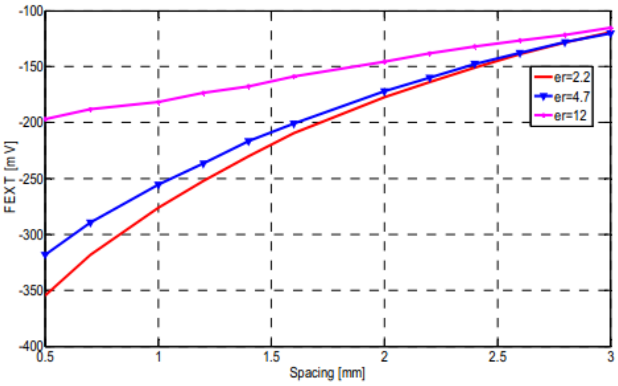


Fig. 10 The FEXT for different dielectric permittivity.

can see that the influence of the dielectric permittivity is not linear. We observe that the crosstalk decreases as the dielectric permittivity decreases, and the dielectric that has a permittivity of 4.7 is a good compromise for the crosstalk.

4.4 Influence of Electrical Signal Dynamic

Another important factor to take into account is the dynamic of the input electrical signal. We studied the influence of the time variation of an input signal depending on the spacing between two microstrip. The calculation per unit length parameters was applied by setting the width of the strip at W = 0.4 mm, for different spacing values and for three configurations of the rise time of input signal: $t_{r1} = 5$ ns, $t_{r2} = 10$ ns, and $t_{r3} = 15$ ns.

For each model, we have taken the peaks of voltages for the NEXT and the FEXT. Figs. 11 and 12 illustrate the variation of the NEXT and the FEXT. We can see that the influence of the rise time is not

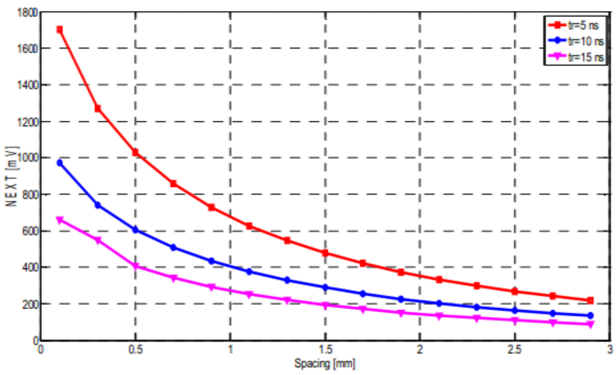


Fig. 11 The NEXT for different rise time.

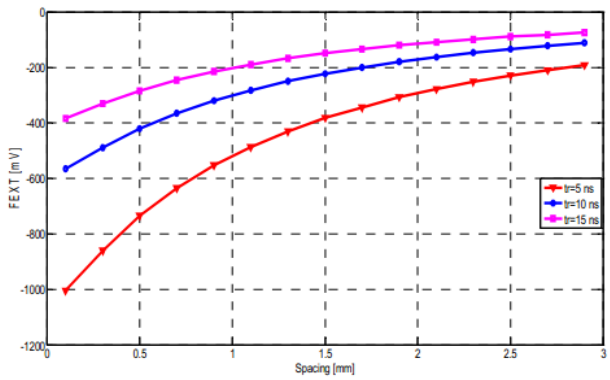


Fig. 12 The FEXT for different rise time.

linear. For example, for a spacing of 0.5 mm, the NEXT changes from 1 V at t_r = 5 ns to 0.6 V at t_r = 10 ns. The knowledge of this parameter is important at the moment of conception.

5. Conclusions

In this work, electromagnetic coupling modeling of coupled microstrip lines used in switching power supplies is presented. The study of the crosstalk requires knowledge of per unit length parameters, of coupled transmission lines. The extraction of per unit length parameters is performed by the method of moments.

To model the electromagnetic couplings between coupled transmission lines, we developed a SPICE model based on electrical parameters of the per unit length. This model allowed us to calculate the NEXT and the FEXT present between coupled transmissions lines constituting a control bus used in switching power supplies.

The crosstalk collected can expect almost 1 V, and -560 mV, respectively, for the NEXT and the FEXT, which can be very harmful. In our application, by increasing the spacing of 0.5 mm, we can divide by two the crosstalk.

By adjusting the physical parameters, and the dynamic of the input signal, we can reduce the influence of the crosstalk by more than 90%, and it becomes less problematic for components and systems close to the control bus.

This work will create curves that will help the designer in the choice of strip width and spacing as a function of electrical stress in the accepted specifications.

References

[1] A. Charoy, Electromagnetic Compatibility, Dunod, France, 2012.

[2] R. Clayton, Introduction to Electromagnetic Compatibility, 2nd ed., John Wiley & Sons, USA, 2006.

[3] A. Boyer, S.A. Boulingui, S. Bendhia, E. Sicard, S. Baffreau, A methodology for predicting by near field chip to chip coupling, in: EMC Compo 2007 (6th International Workshop on Electromagnetic Compatibility of Integrated Circuits), Torino, Italy, Nov.28-30, 2007, pp. 301-306.

[4] S. Musa, M. Sadiku, J. Clark, Finite element analysis for electromagnetic parameters of multiconductor interconnects in multilayered dielectric media, IJAAS (International Journal of Advances in Applied Sciences), 1 (4) (2012) 181-190.

[5] W. Gibson, The Method of Moments in Electromagnetics, Taylor & Francis Group, USA, 2006.

[6] N. Tidjani, J. Le Bunetel, A. Ouchar, Crosstalk analyses in coupled transmission lines used in telephony mobile, in: 17th IEEE Signal and Power Integrity, Paris, 2013, pp. 155-158.

[7] N. Tidjani, J. Le Bunetel, A. Ouchar, Crosstalk estimation in asymmetrical printed circuit boards traces used in telecommunications systems, in: International Conference on Wireless and Mobile Communication Systems (WMCS'13), WSEAS, Paris, France, Oct. 29-31, 2013.

[8] N. Tidjani, J.C. Le Bunetel, A. Ouchar, Y. Raingeaud, An approach to electromagnetic compatibility design rules for control bus used in power electronics, in: ICEEAC (International Conference on Electrical Engineering and Automatic Control), Setif, Algeria, Nov.24-26, 2013.

[9] F. Gardiol, Electricity Treaty: Electromagnetism, Vol. 3, Polytechnic and University, Romande, France, 2004.

[10] J. Pérez, R. Carles, R. Fleckinger, Electromagnetism: Foundations and Applications, Dunod, France, 2001.

[11] R. Clayton, Analysis of Multiconductor Transmission Lines, John Wiley & Sons, IEEE Press, USA , 2008.

Journal of Control Science and Engineering
Volume 1, Number 1, December 2013
David Publishing Company
240 Nagle Avenue #15C, New York, NY 10034, USA
Tel: 1-323-984-7526, 323-410-1082; Fax: 1-323-984-7374, 323-908-0457
http://www.davidpublishing.com
cse@davidpublishing.com, cse@davidpublishing.org

DAVID PUBLISHING

ISSN 2328-2231
12
9 772328 223133